

**Entwicklung und Evaluierung eines Milchprotein-Transkript-
Depletionsverfahrens für differentielle Transkriptomanalyse in
bovinem Milchdrüsengewebe**

von Johanna Brodhagen

Inaugural-Dissertation zur Erlangung der Doktorwürde der
Tierärztlichen Fakultät der Ludwig-Maximilians-Universität
München

**Entwicklung und Evaluierung eines Milchprotein-Transkript-
Depletionsverfahrens für differentielle Transkriptomanalyse in
bovinem Milchdrüsengewebe**

von Johanna Brodhagen
aus Ribnitz-Damgarten

München 2021

**Aus dem Zentrum für Klinische Tiermedizin der Tierärztlichen Fakultät
der Ludwig-Maximilians-Universität München**

Lehrstuhl für Physiologie und Pathologie der Fortpflanzung

Arbeit angefertigt unter der Leitung von: Prof. Dr. Holm Zerbe

**Angefertigt am Leibniz-Institut für Nutztierbiologie (FBN),
Institut für Genombiologie, Dummerstorf**

Mentor: Prof. Dr. Christa Kühn

Gedruckt mit der Genehmigung der Tierärztlichen Fakultät
der Ludwig-Maximilians-Universität München

Dekan: Univ.-Prof. Dr. Reinhard K. Straubinger, Ph.D.

Berichterstatter: Univ.-Prof. Dr. Holm Zerbe

Korreferent/en: Univ.-Prof. Dr. Eckhard Wolf

Tag der Promotion: 6. Februar 2021

Für Arthur

INHALTSVERZEICHNIS

ABKÜRZUNGSVERZEICHNIS	I
ABBILDUNGSVERZEICHNIS	III
1 Einleitung.....	1
2 Erweiterte Literaturrecherche.....	4
2.1 Bovine Mastitis	4
2.2 Biomarker	5
2.3 Next Generation Sequencing.....	5
2.4 RNA-Seq.....	6
2.5 Transkriptom des Eutergewebes.....	7
2.6 Sequenziertiefe.....	9
2.7 Depletion	9
3 Ziel	14
4 Publikation.....	16
5 Diskussion.....	36
5.1 Die Methode.....	36
5.1.1 Entwicklung und Optimierung der Methode	36
5.1.2 Die Auswahl der Versuchstiere	37
5.1.3 Die frühe Laktation als Zeitpunkt der Probennahme	37
5.1.4 Anwendung der Methode an mit <i>E. coli</i> infiziertem und nicht-infiziertem bovinem Eutergewebe	38
5.1.5 Enzymatischer Abbau hoch exprimierter Transkripte mittels RNase H	38
5.1.6 Datenanalyse und Referenzannotation.....	40
5.2 Ergebnisse der Evaluierung der Methode	41
5.2.1 Depletion hochabundanter Milchproteingen-Transkripte aus dem Eutertranskriptom	41
5.2.2 Differenzielle Transkriptomanalyse nach <i>E. coli</i> -Infektion von depletierten und nicht- depletierten Milchdrüsengewebe	43

5.2.3	Identifizierung von Transkripten mit niedrigerem Expressionsniveau durch sensitivere RNA-Seq-Analyse nach der Depletion	47
5.2.4	Einsatz der verbesserten Empfindlichkeit der RNA-Seq-Analyse zur Detektion noch unbekannter Loci.....	48
5.2.5	Vergleich der experimentellen Varianten A und B in Bezug auf Depletionseffektivität, Verzerrung der Ergebnisse und Verbesserung der Sensitivität der RNA-Seq-Analyse	48
5.3	Anwendung der Methode zur Identifizierung vorteilhafter Genotypen.....	49
6	Schlussfolgerung und Ausblick	53
7	Zusammenfassung.....	54
8	Summary.....	57
9	Literatur	59
10	Danksagung	69

ABKÜRZUNGSVERZEICHNIS

A	Adenin
Cas9	CRISPR associated protein 9
CCL	Chemokin Ligand
CCR	Chemokinrezeptor
cDNA	Complementary DNA
CRISPR	Clustered regularly interspaced short palindromic repeats
CXCL	Chemokin
DNA	Desoxyribonukleinsäure
DNase	Desoxyribonuklease
<i>E. coli</i>	<i>Escherichia coli</i>
FPKM	Fragments Per Kilobase per Million reads
GLYCAM	Glycosylation-dependent cell adhesion molecule
GRO	Melanoma growth stimulating activity
ICAM	Intercellular adhesion molecule
IGV	Integrative Genomics Viewer
IL	Interleukin
LALBA	α -Lactalbumin
lncRNA	Long non-coding RNA
miRNA	MicroRNA
MMP	Matrix-Metallopeptidase
mRNA	Messenger RNA
NCBI	National Center for Biotechnology Information
NGS	Next-Generation Sequencing
PAEP	Progesterone-associated endometrial protein
PCR	Polymerase Chain Reaction
QTL	Quantitative Trait Locus
RNA	Ribonukleinsäure
RNase H	Ribonuklease H
RNA-Seq	RNA-Sequencing
ROS	Reactive Oxygen Species
RPKM	Reads Per Kilobase per Million reads
rRNA	Ribosomale RNA
RT-qPCR	Real-time quantitative PCR
sgRNA	Single Guide RNA
SNP	Single Nucleotide Polymorphism
Spp.	Subspecies

S. aureus

T

TLR

TREM1

tRNA

Staphylococcus aureus

Thymin

Toll-like-Rezeptor

Triggering receptor expressed on myeloid cells 1

Transfer RNA

ABBILDUNGSVERZEICHNIS

Abbildung 1:

Methode zur Depletion von hoch abundanten Milchproteingen-Transkripten 13

Abbildung 2:

Entstehung verschiedener Varianten von mRNA durch alternatives Spleißen
von primärer RNA und daraus resultierende Proteinprodukte durch die Translation 43

Abbildung 3:

Differentielle Transkriptomanalyse von Euterparenchym ohne die Anwendung
der Depletion und mit Anwendung der Depletion nach Variante A oder B 44

1 Einleitung

Landwirte beklagen hohe finanzielle Einbußen durch Behandlungskosten, Leistungsabfall, Tierverluste und Kosten für Präventionsmaßnahmen, verursacht durch Mastitiden, welche mit hoher Inzidenz in den Tierbeständen der Milchviehbetriebe auftreten (SEEGERS et al., 2003; HUIJPS et al., 2008; PIEPERS et al., 2009; ROLLIN et al., 2015; VAN SOEST et al., 2016). Mastitis bezeichnet eine Entzündung des Euters mit meist bakteriellem Ursprung und zählt aktuell zu den am häufigsten auftretenden Krankheiten bei Milchkühen (HALASA et al., 2007; FÁVERO et al., 2015). Die Symptome reichen von leichten lokalen Entzündungsanzeichen im betroffenen Euterviertel bis hin zu systemischen Auswirkungen, welche nicht selten zum Tierverlust führen (HOGAN & LARRY SMITH, 2003). Es ist bekannt, dass es sich um eine Faktorenkrankheit handelt, bei welcher der Wirt, wie auch die Umgebung, einen maßgeblichen Einfluss auf die Infektionsrate und die Pathogenese haben (BURVENICH et al., 2003). Die Anfälligkeit gegenüber Euterentzündungen hängt nicht nur vom Alter, Ernährungszustand oder Entwicklungsstatus der Milchkuh ab, sondern wird grundlegend von einem genetisch vorteilhaften oder unvorteilhaften Status geprägt (MEYERHOLZ et al., 2018; PARKER GADDIS et al., 2020).

Um nicht nur eine optimale Behandlung zu gewährleisten, sondern Mastitis möglichst durch genetische Selektion in den Milchviehbeständen zu minimieren, müssen krankheitsassoziierte Veränderungen durch Erkenntnisse über physiologische und pathologische Mechanismen auf molekularer Ebene des Milchkuhorganismus aufgedeckt werden. Außerdem müssen Genomregionen mit genetischen Varianten identifiziert werden, welche die Empfänglichkeit für Mastitis beeinflussen. Die Etablierung stabiler Biomarker als Indikatoren für pathologische bzw. physiologische Merkmale, die Prozesse der Mastitis beeinflussen und die als indirekte Parameter im Zusammenhang mit der Empfänglichkeit gegenüber Euterentzündungen stehen, wäre erstrebenswert. Die Identifizierung vorteilhafter oder unvorteilhafter Genotypen in Bezug auf Mastitisempfindlichkeit wäre so schnell und einfach möglich und diese Biomarker könnten als Werkzeug für die Zucht widerstandsfähiger Milchviehpopulationen eingesetzt werden. Auch eine Anwendung der Biomarker als prognostische Indikatoren für Tiere mit erhöhten Prophylaxe- bzw. spezifischem Therapiebedarf wäre möglich.

Als RNA-Sequencing (RNA-Seq) oder RNA-Sequenzierung bezeichnet man die Bestimmung der Nukleotidabfolge der RNA, basierend auf Hochdurchsatzmethoden der nächsten Generation (Next-Generation Sequencing, NGS) (WANG et al., 2009). Die RNA-Seq hat die Untersuchung ganzer Transkriptome von Zellen und Geweben revolutioniert und bietet grundlegende neue Einsichten in die strukturelle Organisation und funktionelle Regulation der Genome auf Transkriptionsebene, wodurch Einblicke in die globale Genexpression ermöglicht werden (WANG et al., 2009). Eingesetzt als Instrument zur molekularen Katalogisierung zell- und gewebespezifischer Elemente des Transkriptoms dient sie auch der Identifizierung und der funktionellen Annotation regulatorischer Einheiten. RNA-Seq soll dabei helfen, Netzwerke oder Gencluster im Eutergewebe zu identifizieren, welche an den Prozessen der Regulation im Euter beteiligt sind und die Mastitisanfälligkeit beeinflussen.

Eine besonders hohe Inzidenz an Mastitis zu erkranken besteht direkt nach der Abkalbung, d.h. in der frühen Laktation, durch die Geburt und die damit einhergehende hormonelle Umstellung und Stoffwechselveränderungen (VANGROENWEGHE et al., 2005; BURVENICH et al., 2007). Jedoch sind die molekularen Regulationsmechanismen, die für eine unterschiedliche Krankheitsanfälligkeit ursächlich sind, äußerst komplex und noch nicht im Detail geklärt.

Nach der Abkalbung steht die Laktation im Vordergrund des Stoffwechsels der Milchdrüse. Es kommt zur Vermehrung des Euterparenchyms und damit verbunden besteht das Eutertranskriptom in dieser Zeit bis zu 80% aus Transkripten von Genen, welche für Milchproteine kodieren (IBEAGHA-AWEMU et al., 2016). Das Erstellen eines umfassenden Transkriptkataloges des Eutergewebes wird durch diesen Anteil hoch exprimierter Transkripte von Milchproteingenen, zu denen Casein- (CSN2, CSN3, CSN1S1, CSN1S2) und Molkenproteingene (PAEP und LALBA) gehören, beeinträchtigt (WICKRAMASINGHE et al., 2012). Zum Verstehen regulatorischer Vorgänge und zur Identifizierung struktureller Besonderheiten auf molekularer Ebene im Milchdrüsengewebe, welche unterschiedliche Empfänglichkeit gegenüber Mastitis verursachen, ist es jedoch notwendig, auch marginal exprimierte Transkripte und kurze RNAs präzise identifizieren und beurteilen zu können. Es ist bekannt, dass auch Transkripte mit niedrigerem Expressionsniveau, wie häufig bei langen nicht-kodierenden RNAs (lncRNAs) oder kurzen RNAs, wie MicroRNAs (miRNAs) auftretend, regulatorische Funktionen erfüllen, weshalb auch eine Erstellung von Transkriptprofilen mit einer detaillierten Abbildung dieser Sequenzen notwendig ist (LI et al., 2012; DO et al., 2017).

Um trotz der extrem hoch exprimierten Milchproteingen-Transkripte das Auffinden gering vorkommender Sequenzen zuverlässig zu ermöglichen, ist eine hohe Sequenztiefe im Rahmen der RNAseq notwendig, was erhöhte Kosten für den Sequenzierungsvorgang zur Folge hat. Um diese Mehrkosten zu umgehen und trotzdem eine zuverlässige und präzise Identifizierung für die Regulationsvorgänge bedeutender Moleküle bei gegebener Sequenztiefe zu gewährleisten, muss der Anteil der hoch exprimierten Transkripte reduziert werden.

Dieser Ansatz wurde im Rahmen der vorliegenden Dissertation verfolgt. Es sollte ein Verfahren entwickelt und evaluiert werden, welches zur Reduktion der hochfrequenten Milchproteingen-Transkripte im bovinen Eutertranskriptom führt und eine exakte Zell- und Gewebekatalogisierung von gering und hoch exprimierten Transkripten mittels RNA-Seq bei niedrigem Kostenaufwand ermöglicht.

Aus der Literatur und älteren Studien sind bereits Techniken zur Reduzierung von stark exprimierter ribosomaler, mitochondrialer und von Globingenen abgelesener RNA aus Gesamt-RNA zur Verbesserung der Empfindlichkeit der RNA-Seq gegenüber marginal exprimierten Transkripten unterschiedlicher Gewebe bekannt (O'NEIL et al., 2013; HARRINGTON et al., 2020). Die kommerziell genutzten Kits beruhen dabei auf zwei hauptsächlich eingesetzten Methoden. Beide beginnen mit der Hybridisierung spezifischer komplementärer DNA-Oligonukleotide an die gewünschten Zielsequenzen des RNA-Pools. Im zweiten Schritt werden die hybridisierten RNA:DNA-Fragmente dann entweder mittels

paramagnetischer Beads gebunden und extrahiert oder durch eine doppelstrangspezifische RNase H enzymatisch abgebaut. Letztere Methode wird als effektivste und kostengünstigste beschrieben und dient darum in der vorliegenden Studie als Depletionsansatz (ADICONIS et al., 2013).

2 Erweiterte Literaturrecherche

2.1 Bovine Mastitis

Von Mastitis spricht man im Falle einer Entzündung der gesamten Milchdrüse mit ihren milchbildenden, speichernden und ableitenden Abschnitten. Es werden akute und chronische, sowie klinische und subklinische Verläufe beobachtet (SCHULZ, 1994). Gleichzeitig wird Mastitis als infektiöse Faktorenkrankheit mit meist bakteriellem Ursprung beschrieben (ZADOKS et al., 2011; FÁVERO et al., 2015). Als am häufigsten auftretende Krankheit in Milchviehbetrieben (HALASA et al., 2007; BARKEMA et al., 2009; SAINI et al., 2012; YANG et al., 2016), verbunden mit großen wirtschaftlichen Einbußen durch sinkende Milchleistung und Milchqualität, steigende tierärztliche Behandlungs- und Präventionskosten und nicht zuletzt Tierverluste, sind Euterentzündungen von großer Bedeutung (BURVENICH et al., 2003; SEEGERs et al., 2003; HALASA et al., 2007; HUIJPS et al., 2008; DE VliegHER et al., 2012; VAN SOEST et al., 2016).

In der aktuellen Studie von Raboisson et al. (RABOISSON et al., 2020) werden durchschnittliche Kosten von 224€ pro auftretenden Mastitisfall angegeben, was bei einer Prävalenz laut Literatur bei Färsen nach der Abkalbung von bis zu 57% zu erheblichen finanziellen Verlusten der betroffenen Betriebe führt (DE VliegHER et al., 2012).

Ursachen von Euterentzündungen können Bakterien und selten auch Viren, Pilze und Algen sein. Epidemiologisch unterscheidet man Umweltkeime und spezifische Keime, die von Tier zu Tier, hauptsächlich durch den Melkvorgang übertragen werden (CERVINKOVA et al., 2013). Zu diesen übertragbaren, intramammären Krankheitserregern gehören: *Staphylococcus aureus*, *Streptococcus agalactiae*, *Mycoplasma spp.* und *Corynebacterium bovis*. Ubiquitär vorkommende Keime sind *Escherichia coli*, *Klebsiella spp.*, *Streptococcus dysgalactiae* und *Streptococcus uberis* (ABEBE et al., 2016).

Diese Studie wurde an Eutergewebe mit experimentell durch den Umweltkeim *E. coli* induzierter Entzündung und an unbehandeltem Eutergewebe evaluiert. *E. coli* gilt als bedeutendste enterobakterielle Spezies, die an der Ätiologie der klinischen Mastitis beim Rind beteiligt ist und ihren Ursprung in der Umwelt hat. Die durch dieses Bakterium induzierte Infektion der Milchdrüse weist ein breites Spektrum an Symptomen auf, die von lediglich der Änderung der Milchbeschaffenheit, über lokale Entzündungen mit Gefahr des Verlustes des betroffenen Euterviertels, bis hin zu systemischen Krankheitsanzeichen reichen können. Nicht selten ist der Tod des Tieres die Folgedieser Krankheit (HOGAN & LARRY SMITH, 2003).

Die Inzidenz einer klinischen Mastitis im peri- und postpartalen Zeitraum ist für die Milchkuh besonders hoch. Die hormonelle und metabolische Umstellung durch die gerade vollendete Trächtigkeit und Abkalbung machen diesen Zeitraum besonders kritisch und schwächen das Immunsystem (OLIVER & SORDILLO, 1988; BURVENICH et al., 2003; VANGROENWEGHE et al., 2005; BURVENICH et al., 2007). Das in der vorliegenden Studie verwendete Milchdrüsenparenchym wurde in dieser kritischen Phase gewonnen.

Der Schweregrad der bovinen *E. coli*-Mastitis wird hauptsächlich durch Faktoren des Wirts und weniger durch die Pathogenität des eindringenden Erregers bestimmt (BURVENICH et al., 2003). Während der Mastitis ist der Abwehrstatus des Wirts ein wesentlicher Faktor, der den Ausgang und die Pathogenese der Erkrankung beeinflusst. Dieser Status wiederum wird beispielsweise durch die Geburt, das Laktationsstadium, das Alter, die Rasse, den Stoffwechsel, den Ernährungszustand, aktuelles Wachstum und pro- und antiinflammatorische endogene Substanzen geprägt (BURVENICH et al., 2003). Erheblichen Einfluss auf die Empfänglichkeit eines Tieres gegenüber Euterentzündungen hat die Genetik. Aktuelle Studien decken im Genom Regionen mit genetischen Variationen auf, welche dem entsprechenden Tier einen Vorteil in Bezug auf die Krankheitsanfälligkeit verschaffen (MEYERHOLZ et al., 2018; CAI et al., 2020; FREEBERN et al., 2020). Genau hier liegt der Schwerpunkt zukünftiger wissenschaftlicher Studien, mit dem Ziel, eine Optimierung der nächsten Generation von Milchkühen in Hinblick auf Widerstandsfähigkeit und Resistenz gegenüber Krankheiten zu erreichen. Um die genetische Selektion von Milchrindern mit verringerter Empfindlichkeit gegenüber Mastitis zu erleichtern und voranzutreiben, wäre eine Identifizierung zuverlässiger Biomarker von großem Vorteil.

2.2 Biomarker

Biomarker sind messbare, charakteristische, biologische Merkmale, die einen klinisch relevanten Ausgang eines Prozesses oder einen Status, der auf andere Weise nur mit höherem Aufwand festzustellen ist, ersetzen und im Idealfall die Vorhersage über zukünftige Ausprägung von Phänotypen erlauben. Dabei sollte das Auffinden des Biomarkers kostengünstiger, schneller und einfacher sein, als die Detektion des eigentlich für die Studie relevanten Zustandes. Zur Etablierung eines Biomarkers ist das Verständnis über die Beziehung zwischen Marker und dem fokussierten relevanten Status höchst bedeutend (ARONSON & FERNER, 2017). Um durch zukünftige Studien und Forschung stabile Biomarker zur prognostischen Untersuchung der Empfänglichkeit von Milchrindern gegenüber Mastitiden identifizieren zu können, ist es notwendig, die Vorgänge auf molekularer Ebene im Milchdrüsenparenchym vor und nach Pathogenkontakt nachzuvollziehen.

2.3 Next Generation Sequencing

Wissenschaftler nutzen neueste Technik, wie Next Generation Sequencing, für holistische Analysen, um einen Überblick über den Hintergrund und den Zusammenhang von molekularen Mechanismen auf DNA- und RNA-Ebene zu gewinnen. Es ist möglich, diese Technik zur umfassenden Betrachtung physiologischer und pathologischer Vorgänge und unterschiedlicher Krankheitsempfänglichkeit, wie auch Krankheitsverläufe und -ausgang z.B. in Bezug auf Mastitis, zu verwenden.

Next Generation Sequencing (NGS) beschreibt eine Hochdurchsatz-Sequenzierungstechnologie, bei der die Nukleotidabfolge von Millionen von DNA- oder RNA-Fragmenten parallel bestimmt wird. Mithilfe von bioinformatischen Analysen werden diese Fragmente durch das Abgleichen

mit einem Referenzgenom zusammengefügt. Die Sequenzierung des gesamten Genoms ist im Gegensatz zu vorherigen Methoden, bei denen dafür Jahrzehnte benötigt wurden, innerhalb weniger Tage möglich. NGS kann auf kleinere Bereiche oder einzelne Gene des Genoms beschränkt werden (BEHJATI & TARPEY, 2013).

2.4 RNA-Seq

NGS angewendet am Transkriptom wird als RNA-Sequenzierung (RNA-Seq) bezeichnet. Diese Technik wurde in dieser Studie genutzt und dient der quantitativen und sequenziellen Bestimmung der gesamten transkribierten RNA (mRNA, rRNA, tRNA und nicht-kodierende RNA) einer Zelle oder eines Gewebes zu einem bestimmten Zeitpunkt unter gegebenen physiologischen Einflüssen. Sie liefert uns grundlegende Erkenntnisse über strukturelle Organisation und funktionelle Regulation des Genoms. RNA-Seq trägt zum Verstehen von physiologischen Veränderungen in Organismen als Antwort auf sich ändernde Umwelteinflüsse oder krankheitsassoziierte molekulare Mechanismen bei (WANG et al., 2009; OZSOLAK & MILOS, 2011).

Das allgemeine Protokoll der RNA-Seq unter Nutzung der Sequencing by Synthesis Technologie auf Illumina-Plattformen (Illumina, San Diego/USA) beginnt mit der Herstellung einer cDNA-Bibliothek durch Umwandlung transkribierter RNA in cDNA-Fragmente. An die Enden der cDNA-Fragmente werden spezifische Adapter-Oligonukleotide mit funktionellen Elementen ligiert. Mithilfe dieser für eine Plattform (z.B. eine Flowcell oder ein Chip) spezifischen Adapter werden die cDNA-Fragmente an eine feste Oberfläche gebunden und durch Gegenstrangsynthese vervielfältigt. Es entstehen Cluster aus identischen DNA-Molekülen, welche anschließend für eine Vielzahl parallel ablaufender Sequenzierungen je nach gewünschter Zyklenzahl genutzt werden. Die Sequenzierung mittels Sequencing by Synthesis unter Einbau von Fluoreszenzfarbstoff-markierten Nukleotiden erfolgt entweder von nur einem Ende der DNA-Fragmente, nach dem einfachen, günstigen und schnellen Prinzip der single-end-Sequenzierung, oder in beide Leserichtungen, nach dem Prinzip der paired-end-Sequenzierung. Die letztere Variante liefert, bei gleichem Zeit- und Arbeitsaufwand für die Vorbereitung der Bibliotheken, die doppelte Anzahl von Lesevorgängen und vereinfacht das Auffinden von Spleiß-Varianten (Abbildung 2) und neuartigen Transkripten im Genom. Durch das Erkennen von Überlappungen, repetitiven Sequenzelementen oder Inserts (nicht-sequenzierte DNA-Teile) erhöht die paired-end RNA-Seq die Genauigkeit der Sequenzierung und erleichtert bioinformatische Auswertungen signifikant (HOLT & JONES, 2008; WANG et al., 2009; WILHELM & LANDRY, 2009). Die Präparation der cDNA-Bibliotheken kann strangspezifisch oder nicht-strangspezifisch erfolgen. Die Information, welcher DNA-Strang im Genom (Watson- oder Crick-Strang) transkribiert wurde, bleibt bei der ersten Variante erhalten und kann in nachfolgenden bioinformatischen Auswertungen genutzt werden.

Die am Ende des Protokolls millionenfach abgelesenen Reads können an einem Referenzgenom ausgerichtet und zu einer Transkriptom überspannenden RNA-Sequenzkarte zusammengestellt

oder alternativ ohne Zuordnung zu einem Referenzgenom *de novo* zusammengestellt werden (ZHAO et al., 2015).

RNA-Seq bringt im Gegensatz zu früher genutzten Technologien wie Microarrays, die auf Hybridisierung basierten, den Vorteil, auch unbekannte genomische Sequenzen mit neuartigen Transkripten abbilden zu können. Dieses erlaubt auch die Anwendung bei Nicht-Modellorganismen, die oft keine umfassende Transkriptom-Annotation oder ein Referenzgenom aufweisen. Das Auffinden von differentiellen Sequenzvarianten, wie zum Beispiel Einzelnukleotid-Polymorphismen (SNPs) (KANG et al., 2016), alternativer Spleiß-Varianten (STAIGER & BROWN, 2013), der Übergänge von Intron zu Exon, nicht-kodierender RNA-Fragmente (CLARK et al., 2015; ESPOSTI et al., 2016; NIKNAFS et al., 2016) und neuer Transkripte (HALVARDSON et al., 2013), sind möglich (LISTER et al., 2009; WANG et al., 2009). Hintergrundsignale durch Kreuzhybridisierung bleiben aus und der Nachweis von Genen mit sehr hohem oder sehr niedrigem Expressionsniveau ist durch den weiten dynamischen Bereich der Technik möglich. Dabei sind die Ergebnisse zuverlässig reproduzierbar (WANG et al., 2009).

Bei der Anwendung der RNA-Seq-Technik in Verbindung mit Eutergewebe im Zeitraum der Laktation gilt die Berücksichtigung einiger Besonderheiten, auf die im Folgenden eingegangen wird und die die Weiterentwicklung der Methode, wie in dieser Studie geschehen, notwendig macht.

2.5 Transkriptom des Eutergewebes

Während der Laktation weist das Transkriptom der Milchdrüse einen überwiegenden Anteil an Transkripten von Genen kodierend für Milchproteine auf. Dazu gehören Transkripte für Casein-(CN) und Molkenproteine. Die für Caseine kodierenden Gene (CSN1S1, CSN1S2, CSN2 und CSN3) liegen beim Rind auf dem Chromosom 6 (WICKRAMASINGHE et al., 2012). Das Gestagen-assoziierte Endometrialprotein-Gen (PAEP), welches für β -Lactoglobulin kodiert und α -Lactalbumin (LALBA) sind die wichtigsten Molkenproteingene (WICKRAMASINGHE et al., 2012).

Die genannten Milchproteingen-Transkripte werden als höchst exprimierte Gene im Transkriptom des bovinen Euters beschrieben (CANOVAS et al., 2014; IBEAGHA-AWEMU et al., 2016; BHAT et al., 2019). In einigen Veröffentlichungen wurde das Membranprotein der Milchfettkügelchen GLYCAM 1 (Glykosylierungs-abhängiges Zelladhäsions-Molekül 1) noch in die Gruppe der hoch exprimierten Transkripte aufgenommen (CANOVAS et al., 2014; IBEAGHA-AWEMU et al., 2016). Die beschriebene Gruppe der hoch exprimierten Transkripte umfasste in der Studie von Ibeagha-Awemu *et al.* (IBEAGHA-AWEMU et al., 2016) 79,45% der gesamten Transkripte, wobei CSN2 und CSN1S1 am stärksten exprimiert waren. Außerdem wurden die für Milchproteine kodierenden Gene von Cánovas *et al.* (CANOVAS et al., 2014) auch in RNA Sequenzierungen von somatischen Zellen aus Milch, einzelnen durch Laser mikrodisezierten oder Antikörper isolierten Milchdrüsenepithelzellen aus Gewebe und Fettkügelchen aus Milch

von Rindern der Rasse Jersey und Normande, als am höchsten exprimiert identifiziert. Die Caseingene CSN2 und CSN3 wurden dabei in Gewebe, Zellen und Milchfett in beiden Rassen als am höchsten exprimiert festgestellt. Bhat *et al.* (BHAT *et al.*, 2019) isolierten RNA von Epithelzellen aus Milch zu unterschiedlichen Zeitpunkten der Laktation von Kashmiri und Jersey Kühen und zeigten, dass in früher, mittlerer und auch später Laktation CSN1S, CSN1S2, CSN3, CSN2, PAEP und LALBA am höchsten exprimiert wurden. Wickramasinghe *et al.* (WICKRAMASINGHE *et al.*, 2012) zeigten schon 2012, dass die genannten Casein- und Molkenproteine, zusammen mit GLYCAM 1 die Gruppe der am höchsten exprimierten Transkripte in Zellen aus Milch von Holstein Kühen am Tag 15 der Laktation darstellen. Allerdings wurden hier an Tag 90 und 250 der Laktation andere Transkripte als am höchsten exprimiert nachgewiesen.

Auch Studien am ovinen Transkriptom zeigen zu verschiedenen Laktationszeitpunkten bei unterschiedlichen Schafrassen vergleichbare Ergebnisse. 65% der Fragmente wurden hier Milchproteingenen zugeordnet (SUAREZ-VEGA *et al.*, 2016). Die Veröffentlichung von Hao *et al.* (HAO *et al.*, 2019) bestätigt dieses Resultat durch RNA-Seq Analyse von Milchdrüsenparenchym zweier Schafrassen.

Der hohe Anteil transkribierter Milchproteingene in Eutergewebe, den Milchdrüsenzellen und in somatischen Zellen und Fettkügelchen der Milch verhindert oder erschwert das Auffinden marginal exprimierter Transkripte. Da bei der Durchführung der RNA-Seq des Eutertranskriptoms schon bis zu 80% der Lesungen durch Milchproteingen-Transkripte entstehen, verbleibt nur ein geringer Anteil für relevante Transkripte. Um das Auffinden und Identifizieren dieser Transkripte trotzdem präzise zu ermöglichen, ist eine Erhöhung der Sequenziertiefe, also eine höhere Anzahl an Lesevorgängen, nötig. Doch gerade auch Sequenzen mit niedrigerer Expression, wie beispielsweise lncRNAs (im Vergleich zu Protein kodierenden RNAs) oder sehr kurze miRNAs, sind von Bedeutung, was die Regulation von physiologischen und pathologischen Vorgängen im Eutergewebe betrifft.

lncRNAs sind nicht-kodierende RNAs aus mindestens 200 Nukleotiden (PERKEL, 2013) mit Relevanz in Prozessen des Chromatinumbaus, Transkriptionskontrolle und posttranskriptionaler Modifikation. Es besteht ein zunehmendes wissenschaftliches Interesse an der möglichen Beteiligung von non-coding RNAs (ncRNAs) an der Ätiologie von Krankheiten aufgrund gestörter Funktion von ncRNAs bei Differenzierungs- und Entwicklungsprozessen. Die Fähigkeit von ncRNAs, assoziierte Protein kodierende Gene zu regulieren, könnte zur Krankheitsentstehung beitragen, wenn ihre Fehlexpression ein Gen von klinischer Bedeutung fehlerhaft reguliert (MERCER *et al.*, 2009). MicroRNAs sind kurze, für gewöhnlich aus 21 bis 23 Nukleotiden bestehende, hoch konservierte, nicht-kodierende RNAs (LEE *et al.*, 2004). Studien bewiesen schon vor einigen Jahren, dass diese Art von RNAs die Expression von Genen auf transkriptionaler Ebene regulieren (BARTEL, 2004; HE & HANNON, 2004; LAWLESS *et al.*, 2013). In einer aktuelleren Studie deckten Yang *et al.* (YANG *et al.*, 2018) durch RNA-Seq von Milchdrüsengewebe laktierender und trockenstehender Kühe Interaktionen zwischen lncRNAs

(in der Regel mindestens 200 Nukleotide lang) und potentiell zugehörigen miRNAs auf, welche für die Regulation des Laktationszyklus Gewichtung haben. Das wiederum spricht den lncRNAs einen entscheidenden Einfluss bei der Proliferation und Differenzierung von Milchdrüsenzellen zu.

2.6 Sequenziertiefe

Die Nachweisbarkeit eines Transkripts hängt von seiner Länge und Häufigkeit in der zu sequenzierenden Bibliothek ab (SIMS et al., 2014). Um trotz des hohen Anteils der Milchproteingen-Transkripte weniger stark exprimierte Transkripte im Eutergewebe auffinden und präzise identifizieren zu können, ist eine Transkriptom-Sequenzierung mit sehr hoher Sequenziertiefe notwendig. Allgemein definiert die Sequenziertiefe dabei die Anzahl, wie oft ein bestimmtes Nukleotid der Rohsequenz während der Analyse sequenziert wurde (SIMS et al., 2014). Für die RNA-Seq allerdings ist weniger die Anzahl der Sequenzierungen eines Nukleotids als vielmehr die Anzahl von sequenzierten Fragmenten für die Informativität eines Datensatzes entscheidend. Umso mehr Fragmente sequenziert werden müssen, umso höher sind auch die Kosten für diesen Aufwand.

Eine Möglichkeit diesen Mehraufwand und die Kosten zu minimieren, dabei die Qualität und Empfindlichkeit der Detektion von marginal exprimierten Transkripten zu erhalten oder sogar zu verbessern, ist die Anreicherung der für die Studie relevanten RNA-Fragmente oder die Entfernung bzw. Depletion von unerwünschten Fraktionen (SIMS et al., 2014).

2.7 Depletion

Das Prinzip der Anreicherung gewünschter Fragmente oder die Depletion von unerwünschten Transkripten aus dem zu untersuchenden Pool an RNA wurde bereits anderweitig angewendet und wird im Folgenden genauer betrachtet:

Um eine präzise Detektion von mRNA durch RNA-Seq zu gewährleisten ist es beispielsweise notwendig, die in jeder Gesamt-RNA aus Gewebe oder Zellen von mindestens 80-90% vorkommende ribosomale Ribonukleinsäure (rRNA) zu entfernen (O'NEIL et al., 2013). Ebenfalls ist es für viele Spezies vorteilhaft, vor der Sequenzierung peripheren Vollbluts den dominierenden, extrem hoch exprimierten Anteil der RNA-Fragmente des Globingenclusters, welcher 80-90% der Transkripte ausmacht, zu reduzieren, um eine verbesserte Empfindlichkeit der RNA-Seq gegenüber Transkripten niedrigerem Expressionsniveaus im Blut zu erreichen (SHIN et al., 2014; HARRINGTON et al., 2020). Beim Rind allerdings ist dies aufgrund geringer Expressionslevel der Globingene nicht erforderlich (DEMASIUS et al., 2013).

Eine einfache häufig genutzte Methode zur Reduzierung von rRNA-Fragmenten durch die Anreicherung der gewünschten Ziel-Sequenzen in eukaryotischen RNA-Sequenzierungen ist das selektive Primen von mRNAs durch das Hybridisieren von Oligo(dT)nukleotiden an den Poly(A)-Schwanz dieser. Nur die hybridisierten mRNAs werden nachfolgend in cDNA umgeschrieben. Nachteil dieser Methode ist, dass nicht nur eine Entfernung der ribosomalen RNA stattfindet,

sondern jede nicht-polyadenylierte RNA aus der reversen Transkription und somit auch aus der zu erstellenden cDNA-Bibliothek ausgeschlossen wird. Das polyadenylierte, Stabilität gebende und für den nukleozytoplasmatischen Export und die Translation funktionelle 3' Ende entsteht posttranskriptionell an beinahe jeder eukaryotischen mRNA (MOORE & PROUDFOOT, 2009). Ausnahmen bilden unter anderem replikationsabhängige Histone (DETKE et al., 1978), einige lncRNAs (WILUSZ et al., 2008) und die gesamte prokaryotische mRNA (GIANNOUKOS et al., 2012). Zusätzlich gibt es Transkripte, welche einer Gruppe von „bimorphen“ zugeordnet werden. Unter bestimmten Bedingungen wird die Polyadenylierung dieser Moleküle reduziert oder sogar komplett entfernt (KATINAKIS et al., 1980; GU et al., 1999; ZHANG et al., 2014). Um während der RNA-Seq auch diese nicht-polyadenylierten RNAs in die cDNA-Bibliothek aufnehmen zu können, müssen andere Techniken zur gezielten Depletion/Entfernung der unerwünschten Transkripte, wie rRNA oder globine RNA, vorgenommen werden.

Kommerzielle Kits reduzieren hoch exprimierte Globin-RNA aus Blut und ribosomale, wie auch mitochondriale, RNA-Fractionen aus dem Pool der Gesamt-RNA unterschiedlicher Gewebe. Dabei basieren die Kits grundsätzlich auf zwei unterschiedliche Strategien des Abbaus (BUSH et al., 2017). Ausgang ist in jedem Fall die Hybridisierung von spezifischen komplementären DNA-Oligonukleotiden an eine gezielte Gensequenz im RNA-Pool. Die Unterscheidung der Strategien beginnt im zweiten Schritt, während der Eliminierung der unerwünschten Zieltranskripte aus der Gesamt-RNA. Einerseits kommen hier paramagnetische Beads zum Abfangen der RNA:DNA-Hybride zum Einsatz, gebundene RNA wird extrahiert und so aus der Reaktion entfernt, andererseits verwendet man im Ablauf des Protokolls das Enzym RNase H zum Abbau der durch den Doppelstrang entstandenen DNA:RNA-Moleküle, welche im weiteren Verlauf des Protokolls dann nicht mehr zur Verfügung stehen, sofern die eingesetzten Oligonukleotide die gesamte Sequenz der Zieltranskripte abdecken.

Die Depletion durch die genannten Methoden erlaubt einen effizienten Nachweis sowohl funktionell relevanter, kodierender als auch nicht-kodierender, Transkripte (HERBERT et al., 2018). Außerdem ist eine Anwendung auch an RNA niedriger Qualität oder Quantität, wie abgebaute RNA oder RNA geringer Menge aus klinischen Proben, Formalin fixiert oder Paraffin versetzt, gering vorkommenden Zellpopulationen (Sequenzierung einzelner Zellen) oder Kadavern möglich, was für die RNA-Seq oft eine Herausforderung darstellt (ZHAO et al., 2014).

Der Vollständigkeit halber, um einen Überblick existierender Möglichkeit zu erhalten, können noch weitere seltener genutzte Varianten der Depletion von ribosomaler RNA genannt werden: Studien beschreiben das Entfernen der rRNA auf der cDNA-Ebene nach Präamplifikation und Re-Annealing der Bibliothek mittels doppelstrangspezifischer DNase, um die beispielsweise durch Sonden gebundene, von rRNA abgeleitete cDNA zu verdauen (YI et al., 2011; ARCHER et al., 2014; HERBERT et al., 2018).

Des Weiteren beschrieben Castro *et al.* (CASTRO *et al.*, 2013) die Anwendung von Hochleistungs-Flüssigkeitschromatographie als Methode zur Abtrennung prokaryotischer rRNA von Gesamt-RNA. Dabei nutzten sie die unterschiedliche UV-Absorption von mRNA und rRNA.

Hirakawa *et al.* (HIRAKAWA *et al.*, 2011) setzten Hexamere zum Primen der Amplifikation von RNA ein. „Nicht zufällige“ Hexamere, welche identifiziert wurden rRNA zu binden, wurden hier aus dem endgültigen Primer-Pool entfernt, um Sequenzen mit nicht-ribosomalem Ursprung anzureichern.

Eine weitere Technologie, um unerwünschte Transkripte während der Vorbereitung der cDNA-Bibliothek zu minimieren, basiert auf einer Spaltung mittels der CRISPR-assoziierten Nuklease Cas9. Ein Pool aus Single-Guide-RNAs (sgRNAs) steuert dabei den Abbau der von rRNA abgeleiteten cDNA, welche im folgenden PCR Schritt nicht amplifiziert wird (GU *et al.*, 2016; PREZZA *et al.*, 2020).

Zur Depletion hoch exprimierter Milchproteingen-Transkripte aus Milchdrüsenparenchym wurden in dieser Studie DNA-Oligonukleotide, spezifisch für Abschnitte nahe des 3'Endes der Milchproteingen-Transkripte, eingesetzt. Die Hybridisierung dieser Nukleotide an die Zielsequenzen führte zur RNA:DNA-Doppelstrangbildung in diesen Bereichen. Der enzymatische Verdau mittels doppelstrangspezifischer RNase H führte folglich zur Abtrennung des Poly(A)-Schwanzes der hoch exprimierten Zieltranskripte, also zur Deadenylierung. Im nächsten Schritt, der Poly(A)-Selektion, wurden die deadenylierten Milchproteingen-Transkripte aus dem Pool der Gesamt-RNA nicht miterfasst und infolgedessen nicht in die cDNA-Bibliothek eingebaut (Abbildung 1). In Vorversuchen wurden genaue Parameter, wie die Temperatur während der Hybridisierung und die Zusammensetzung der Oligonukleotid-Sets, getestet und festgelegt, mehrere technische Optimierungsschritte durchgeführt und mittels vergleichender RT-qPCR behandelter und unbehandelter RNA analysiert. Um die Wirkungsweise und Effizienz dieser Methode zu evaluieren, erfolgte eine vergleichende RNA-Seq-Analyse. Als Probenmaterial diente dabei mit *E. coli* experimentell infiziertes und unbehandeltes Milchdrüsenparenchym dreier Holstein-Friesian-Erstkalbinnen.

Huang *et al.* (HUANG *et al.*, 2020) beschreiben die Variante mit Anwendung der RNase H als kostengünstige und Adiconis *et al.* (ADICONIS *et al.*, 2013) zusätzlich noch als effektivste für diese Anwendung.

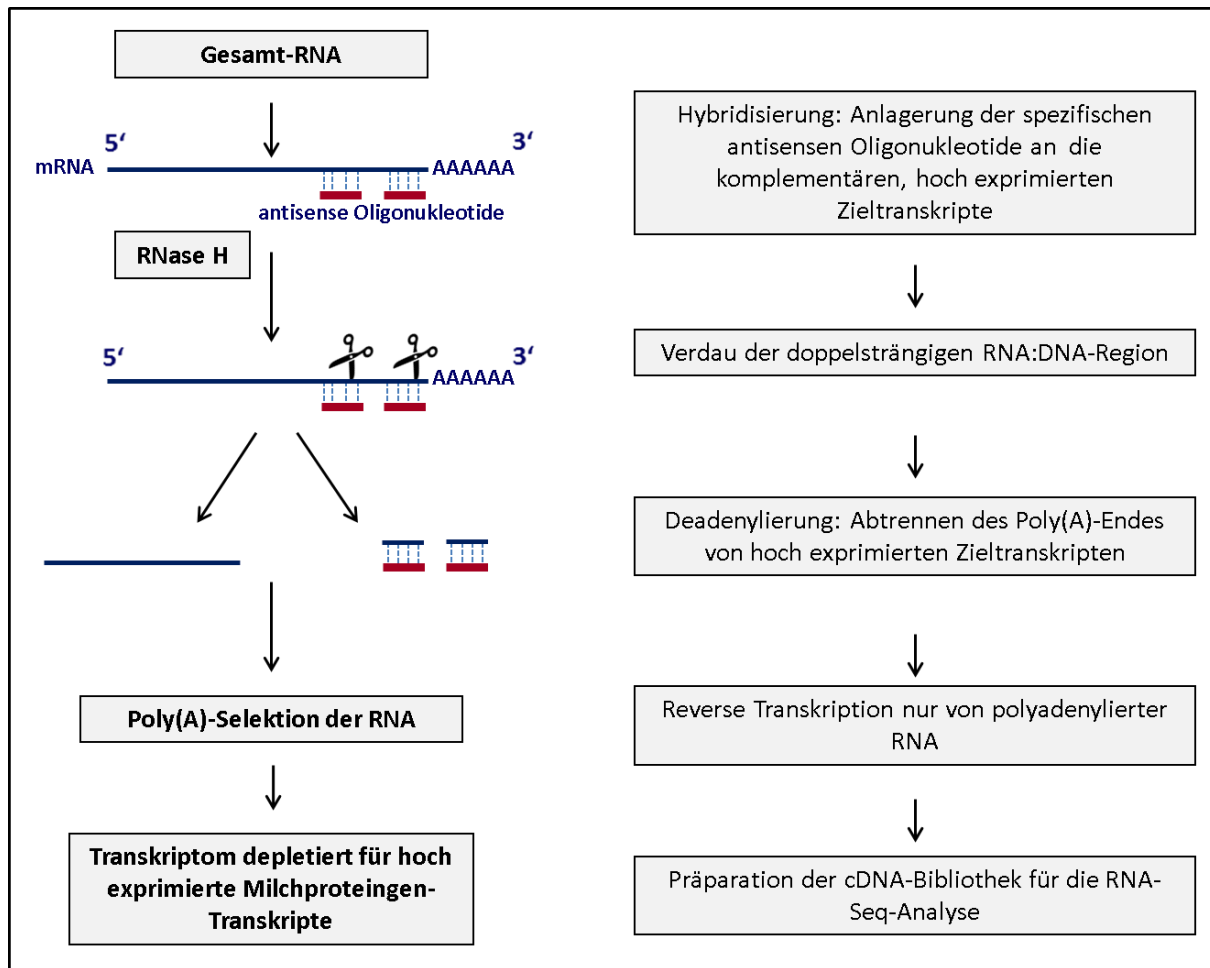


Abbildung 1: Methode zur Depletion von hoch abundanten Milchproteingen-Transkripten

3 Ziel

Ziel dieser Studie war es, eine zuverlässige und kostengünstige Methode zur Entfernung extrem stark exprimierter Transkripte von Genen kodierend für Casein- und Molkenproteine aus dem Pool der Gesamt-RNA von laktierendem bovinem Eutergewebe zu entwickeln, um eine Verbesserung der Sensitivität von RNA-Seq-Analysen gegenüber Transkripten marginaler Expressionsniveaus zu erreichen. Die Entwicklung und Evaluierung dieser Methode soll die Anwendung von RNA-Seq an bovinem Eutergewebe kostengünstiger und präziser ermöglichen und bei gegebener Sequenziertiefe eine detaillierte Katalogisierung zell- und gewebespezifischer Transkriptomelemente erleichtern.

4 Publikation

Johanna Brodhagen, Rosemarie Weikard, Ulrike Thom, Annika Heimes, Juliane Günther, Frieder Hadlich, Holm Zerbe, Wolfgang Petzl, Marie M. Meyerholz, Martina Hoedemaker, Hans-Joachim Schuberth, Susanne Engelmann und Christa Kühn

Development and evaluation of a milk protein transcript depletion method for differential transcriptome analysis in mammary gland tissue

Brodhagen *et al.* BMC Genomics (2019)


20:400 <https://doi.org/10.1186/s12864-019-5781-3>

RESEARCH ARTICLE

Open Access

Development and evaluation of a milk protein transcript depletion method for differential transcriptome analysis in mammary gland tissue



Johanna Brodhagen¹, Rosemarie Weikard¹, Ulrike Thom¹, Annika Heimes¹, Juliane Günther¹, Frieder Hadlich¹, Holm Zerbe², Wolfgang Petzl², Marie M. Meyerholz², Martina Hoedemaker³, Hans-Joachim Schuberth⁴, Susanne Engelmann^{5,6} and Christa Kühn^{1,7*} 

Abstract

Background: In the mammary gland transcriptome of lactating dairy cows genes encoding milk proteins are highly abundant, which can impair the detection of lowly expressed transcripts and can bias the outcome in global transcriptome analyses. Therefore, the aim of this study was to develop and evaluate a method to deplete extremely highly expressed transcripts in mRNA from lactating mammary gland tissue.

Results: Selective RNA depletion was performed by hybridization of antisense oligonucleotides targeting genes encoding the caseins (*CSN1S1*, *CSN1S2*, *CSN2* and *CSN3*) and whey proteins (*LALBA* and *PAEP*) within total RNA followed by RNase H-mediated elimination of the respective transcripts. The effect of the RNA depletion procedure was monitored by RNA sequencing analysis comparing depleted and non-depleted RNA samples from *Escherichia coli* (*E. coli*) challenged and non-challenged udder tissue of lactating cows in a proof of principle experiment. Using RNase H-mediated RNA depletion, the ratio of highly abundant milk protein gene transcripts was reduced in all depleted samples by an average of more than 50% compared to the non-depleted samples. Furthermore, the sensitivity for discovering transcripts with marginal expression levels and transcripts not yet annotated was improved. Finally, the sensitivity to detect significantly differentially expressed transcripts between non-challenged and challenged udder tissue was increased without leading to an inadvertent bias in the pathogen challenge-associated biological signaling pathway patterns.

Conclusions: The implementation of selective RNase H-mediated RNA depletion of milk protein gene transcripts from the mammary gland transcriptome of lactating cows will be highly beneficial to establish comprehensive transcript catalogues of the tissue that better reflects its transcriptome complexity.

Keywords: RNA depletion, RNA-sequencing, Mammary gland, Lactation, Mastitis

* Correspondence: kuehn@fhn-dummerstorf.de

¹Leibniz Institute for Farm Animal Biology (FBN), Institute of Genome Biology, 18196 Dummerstorf, Germany

⁷Agricultural and Environmental Faculty, University Rostock, 18059 Rostock, Germany

Full list of author information is available at the end of the article



© The Author(s). 2019 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

Background

RNA sequencing (RNA-Seq) has revolutionized the study of whole transcriptomes in cells and tissues and has opened a new horizon for the understanding of global gene expression by providing fundamental new insights into the structural organization and functional regulation of the genomes at transcriptional level. Molecular cataloguing of cell- and tissue-specific transcriptome elements is essential for the identification and functional annotation of regulatory features. These are important to understand physiological changes and molecular mechanisms in an organism in response to environmental challenges, during development and under disease-associated conditions.

Bovine mastitis is an inflammatory disease of the mammary gland mainly due to bacterial infection [1, 2]. It affects welfare and health of the cows, resulting in financial losses due to reduced performance, increased treatment costs and animal losses, especially in dairy farming [3–7]. Cows often suffer from mastitis during early lactation, highlighting this period to be important for the investigation of individual mastitis susceptibility [8, 9]. However, the molecular regulatory mechanisms involved in different mastitis susceptibility of cows are complex and not yet clarified in detail. Transcriptomic studies using RNA-Seq can help to identify the gene clusters or networks that are involved in the regulation of processes affecting mastitis susceptibility and incidence in the mammary gland. Transcriptome profiling and the identification of a comprehensive transcript catalogue of the mammary tissue might be compromised due to the fact that at the stage of lactation the transcriptome of the mammary gland is overrepresented by highly abundant transcripts of genes encoding milk proteins, such as genes from the casein gene cluster (*CSN2*, *CSN3*, *CSN1S1*, *CSN1S2*) and whey protein genes, progesterone-associated endometrial protein gene (*PAEP*) encoding β lactoglobulin and α lactalbumin (*LALBA*) [10, 11]. Transcriptome studies in mammary tissue and milk cells of dairy cows have confirmed this transcriptional pattern and reported that the percentage of milk protein gene transcripts can account for up to 70% of all transcripts expressed in this tissue, which may impair the detection of transcripts of genes with a marginal expression level at a given sequencing depth [12]. As a consequence, transcripts expressed at a lower level compared to protein coding genes, e.g., long noncoding RNAs, which may be involved in regulatory processes of immune defence in the lactating udder during infection [13–15], will possibly not be recognized by global transcriptome analysis.

In order to be able to identify even rare transcripts accurately and reliably in the lactating mammary gland transcriptome, a high level of sequencing depth must be

achieved, but this ultimately leads to higher sequencing costs. The aim of our study was therefore to develop a reliable, cost-efficient method that reduces the proportion of high-frequency transcripts in mRNA from bovine mammary gland tissue in order to be able to establish a comprehensive catalogue of mammary gland transcripts. Using this method, an improvement of detection sensitivity of transcripts with marginal expression levels should be achieved while simultaneously the costs for deep transcriptome profiling of mammary gland from lactating cows via RNA-Seq are reduced.

There are two predominant techniques underlying the different RNA depletion procedures and commercially available kits applied to reduce the prevalence of highly abundant genes, such as globin RNA in blood and ribosomal and mitochondrial RNA fractions in total RNA from various tissues. In the first step, both strategies rely on hybridization of gene-specific complementary oligonucleotides to the targeted gene sequences in the total RNA. The difference between these techniques is in eliminating the unwanted targeted genes from the pool of total RNA sequences. In the second step, therefore one strategy is based on capturing of the RNA:DNA duplexes using the paramagnetic bead technology. In the alternative strategy, the RNA:DNA hybrids are degraded by RNase H so that the targeted sequences are no longer available for subsequent applications such as polyA+ selection. A cross-site comparative study with commercially available rRNA depletion kits including kits based on the capture of rRNA by complementary oligonucleotides coupled to beads and also kits based on the hybridization of rRNA with antisense DNA oligonucleotides followed by degradation of RNA:DNA hybrids with RNase H, showed that although there were differences between the underlying rRNA depletion chemistries, all tested kits were able to successfully remove a significant amount of rRNA in library preparations [16]. All kits were able to remove ribosomal RNA to below 20%, but in comparison, the kits that degraded the rRNA by RNase H treatment showed more consistent results than kits that used the bead-based capture method for rRNA depletion. Furthermore, Adiconis et al. [17] and Herbert et al. [16] found that the RNase H-mediated method performed best for rRNA depletion in case of low quality RNA.

In our study we developed an RNase H-mediated RNA depletion approach targeting highly expressed milk protein genes, which was validated experimentally on total RNA isolated from *E. coli* challenged and non-challenged mammary gland tissue of three lactating Holstein-Friesian dairy cows. To monitor the performance and efficiency of this targeted RNA depletion protocol, a comparative RNA-Seq analysis was carried out on depleted and non-depleted mammary gland RNA samples.

Results and discussion

Optimization of the RNase H-mediated RNA depletion procedure for milk protein genes highly prevalent in the mammary gland transcriptome

Targeted RNA removal of highly abundant transcripts prior to RNA-Seq has been successfully introduced into commercial library preparation kits to deplete globin RNA in blood and ribosomal and mitochondrial RNA fractions in total RNA from various tissue types [18–21].

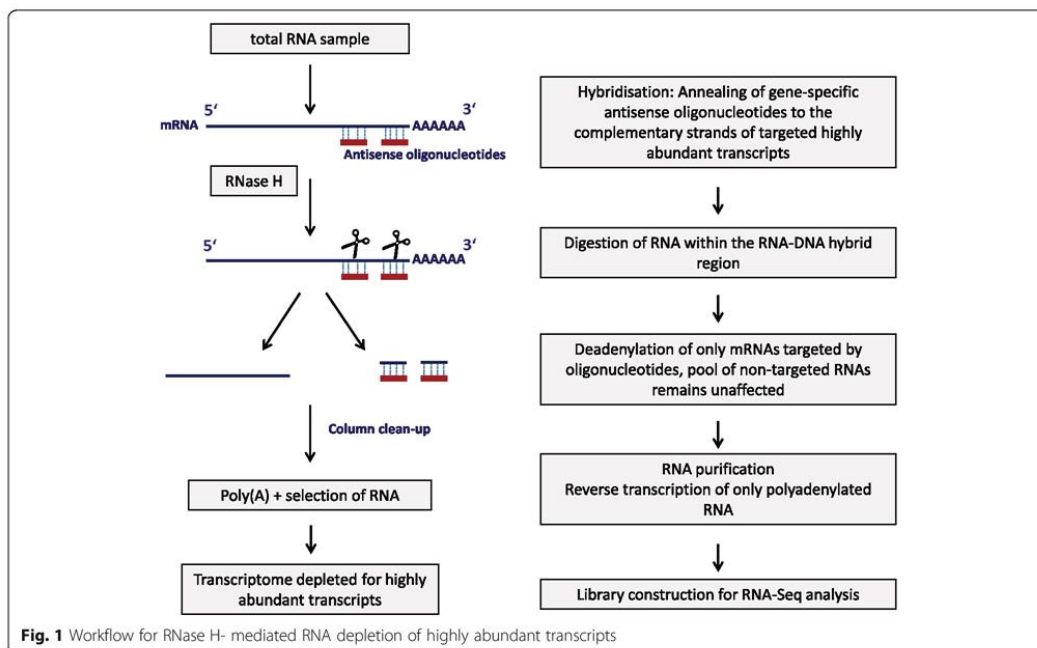
Focusing on the removal of highly abundant milk protein transcripts in the total RNA pool extracted from mammary gland samples of lactating cows, we have developed an RNase H-mediated RNA depletion method based on sequence-specific anti-sense oligonucleotides targeting milk protein genes. The final methodological procedure is illustrated in Fig. 1.

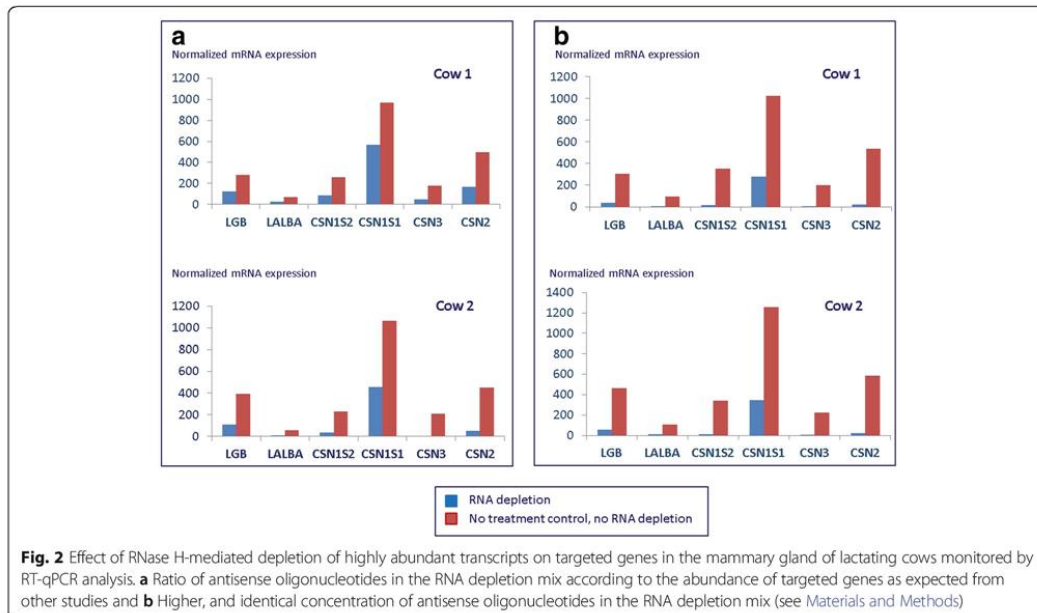
Prior to applying this RNA depletion strategy in a proof of principle experiment on RNA samples from pathogen-challenged and non-challenged udder tissue of lactating cows, several technical optimization steps were performed and monitored by comparative RT-qPCR analysis of depleted and non-depleted RNA samples. Total RNA extracted from udder samples of two lactating cows was used to analyze the effect of modifications in the RNA depletion protocol.

In the initial experiment we adjusted the ratio of anti-sense oligonucleotides in the oligonucleotide depletion mix according to the abundance of the targeted milk

protein genes in lactating mammary gland as retrieved from literature data [10]. Since RT-qPCR monitoring of the depletion treatment showed a smaller depletion effect for *LALBA*, *CSN1S1* and *PAEP* compared to *CSN3*, *CSN2* and *CSN1S2*, we modified the hybridization assay conditions and replaced the antisense oligonucleotide sets for *LALBA*, *CSN1S1* and *PAEP* with others. These modifications established in RNA depletion variant A improved the RNA depletion efficiency only slightly. In addition, the effect of the RNA depletion procedure itself (RNA treatment with buffer instead of antisense oligonucleotides) was also monitored.

Finally, the concentration of the respective oligonucleotides in the depletion mix was modified and adjusted to the same equimolar level for all oligonucleotides for RNA depletion variant B. In the end, the results of RT-qPCR analysis of depleted and non-depleted samples revealed that the mRNA abundance of targeted milk protein genes was reduced by 30 to 90% in the depleted samples depending on the specific gene targeted and the experimental conditions applied. Examples of RT-qPCR evaluation for the two methodological variants A and B used in our follow-up study on RNA samples from pathogen-challenged and non-challenged udder tissue of lactating cows are shown in Fig. 2. Comparing the RT-qPCR results of RNA depleted and non-depleted samples, we see the highest RNA depletion effect in variant B.





Exemplarily, test RNA-Seq libraries were prepared from depleted and non-depleted RNA samples from the same tissue sample and subjected to paired-end RNA-Seq. Monitoring of the RNase H-mediated depletion effect by RNA-Seq analysis revealed a decreased proportion of reads mapping to the targeted milk protein genes, declining from about 60% in the transcriptome for the non-depleted sample to 30% in the RNA depleted sample in non-challenged tissue samples (data not shown). Thus, the results of the transcriptome analysis essentially confirmed those obtained by RT-qPCR indicating a substantial reduction of the prevalence of milk protein genes in the transcriptome of the mammary gland from lactating cows. Nevertheless, we were still able to detect milk protein gene transcripts by both RT-qPCR and RNA-Seq.

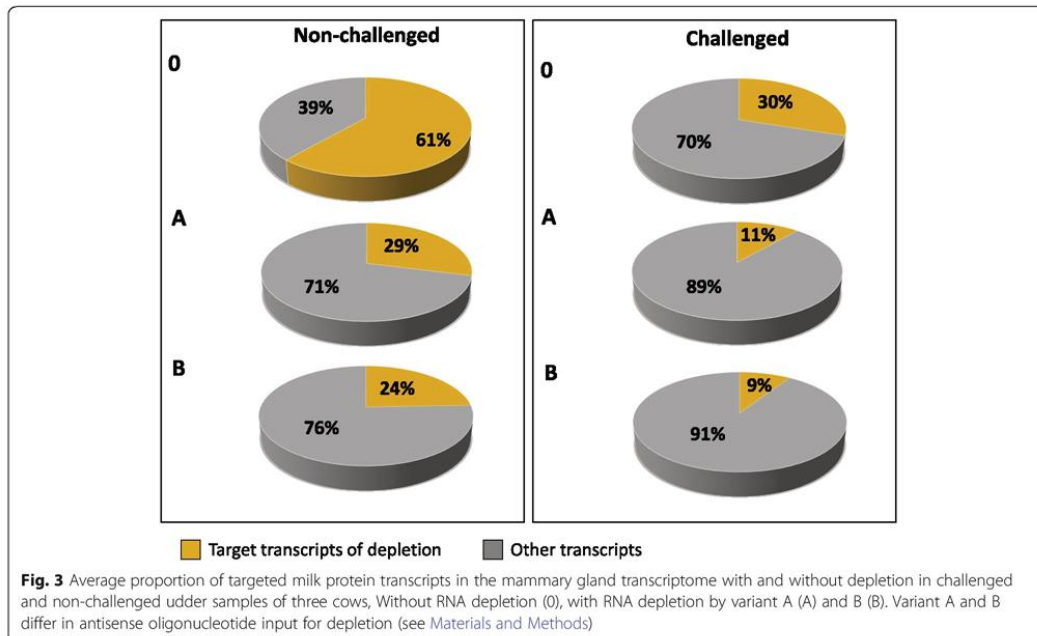
However, the use of an RNA depletion approach, which reduces the total milk protein gene content by about 50% in total, enables sequencing costs to be reduced. We can assume that usually about 60% of the genes expressed in the transcriptome of the lactating mammary gland are milk protein genes (see Fig. 3). For describing the complex transcriptome of this tissue, we need at least 30 mill paired-end non-milk protein gene reads and respectively a sequencing depth of about 75 mill raw reads in non-depleted RNA samples. If we can achieve a 50% reduction of the proportion of milk protein transcripts by applying the selective RNA depletion step (i.e. their proportion in the transcriptome is only

30%), a sequencing depth of 42 mill reads would be sufficient for RNA-depleted samples. This would reduce sequencing costs by 50% per sample (currently 625 to 350 € given standard full cost calculation) at additional costs of 8 € per depletion assay. Thus, for break-even, the costs for sequencing 30 million paired-end non-milk protein gene reads would have to be in the dimension of less than about 50 €. A further advantage of the milk protein depletion is savings in computing time and data storage due to lower number of reads for processing.

Application of the RNaseH-mediated RNA depletion approach in mammary transcriptome sequencing

In a proof-of-principle study on RNA samples from pathogen-challenged and non-challenged udder tissue of lactating cows, transcriptome sequencing analysis has been carried out with a total of 18 RNA-Seq libraries prepared from the mammary gland of three lactating cows. One udder quarter of each cow had been experimentally challenged with *E. coli*. Another udder quarter of the same cow served as a non-challenged control.

RNase H-mediated RNA depletion targeting highly abundant milk protein gene transcripts was carried out on total RNA isolated from challenged and non-challenged mammary gland samples prior to library preparation. RIN values of the non-depleted RNA samples were in a range from 7.7 to 9.6; the scores have declined by about 2 units after RNA depletion treatment, as we had also seen in initial investigations, which may indicate partial RNA



degradation. However, the RIN values were not associated to the RNA depletion differences for individual target genes (see below).

Challenge of udder quarters with the pathogen *E. coli* was performed to compare the effects of the targeted milk protein gene depletion procedure in case of lower milk protein synthesis/milk production and to evaluate a possible bias in the samples caused by applying the RNA depletion method.

Statistics of mammary gland transcriptome sequencing

After adapter trimming of raw reads, 62,510,886 to 92,966,426 reads with a length of 100 bp were obtained for the different mammary gland transcriptome libraries. After read quality trimming, there were between 54,303,133 and 82,880,381 reads available for subsequent transcriptome analysis of depleted and non-depleted RNA samples from challenged and non-challenged udder quarters (Table 1). Guided alignment of these reads to the *Bos taurus* reference genome revealed an average mapping rate of 98% across all 18 samples (Table 1).

Targeted reduction of highly abundant transcripts in mammary transcriptome libraries

Previous transcriptome studies [10, 22, 23] showed that the majority of transcripts expressed in the lactating mammary gland transcriptome mapped to highly abundant

milk protein genes. Ibeagha-Awemu et al. [21] have identified the 24 most frequently expressed genes in the bovine mammary gland transcriptome, with the milk protein genes (*LALBA*, *PAEP*, *CSNIS1*, *CSNIS2*, *CSN2* and *CSN3*) accounting for a total of 77% of reads. The study by Cánovas et al. [10] has shown that the milk protein genes targeted by our RNA depletion method are among the six genes most highly expressed in the mammary gland tissue of lactating Jersey and Normande cows. In the transcriptome of ovine lactating mammary gland tissue, 53% of reads covered transcripts coding for milk protein genes [22].

In our study, RNA-Seq analysis of non-depleted RNA samples from non-challenged udder quarters assigned 61% (52–71%) of the total number of mapped fragments to milk protein genes, whereas in *E. coli* challenged udder quarters 30% (0.6–57%) of fragments mapped to transcripts coding for milk proteins (Fig. 3). Analogous to this high variability, a high clinical variability between cows in response to *E. coli* challenge had also been reported in other studies [24, 25] and in our own dataset (see below).

The lower milk protein gene expression levels in challenged mammary samples might be due to an activation of processes required for immune defense in the mammary gland, which has a higher priority than milk protein gene synthesis under challenge conditions, finally resulting in lower milk yield in pathogen challenged

Table 1 Read statistic of raw data, after quality trimming and read alignment to the bovine genome

Udder quarter treatment	Animal No.	Sample treatment ^a	RIN value	Reads (total number)	Reads after quality trimming (%)	Mapped reads (%)
Non-challenged	Cow1	0	7.7	77,920,096	90	98
		A	4.9	70,975,008	77	97
		B	4.6	81,396,012	91	98
	Cow2	0	7.8	78,859,264	90	98
		A	4.9	80,349,800	90	98
		B	5.4	78,824,666	89	98
	Cow3	0	8.1	80,203,836	90	99
		A	5.1	62,510,886	89	98
		B	5.6	74,101,514	89	98
Challenged	Cow1	0	7.9	78,564,210	90	98
		A	4.7	78,719,102	90	98
		B	4.9	80,990,200	89	98
	Cow2	0	9.6	92,966,426	89	98
		A	5.9	63,610,608	88	95
		B	6.9	70,529,798	89	97
	Cow3	0	7.8	85,090,282	89	98
		A	6.0	81,278,402	89	98
		B	5.7	88,169,812	89	98

^aPathogen challenged and non-challenged udder tissue samples of the cows 1–3, each of them without RNA depletion (0) and with RNA depletion variant A (A) or variant B (B). Variant A and B differ in antisense oligonucleotide input for depletion (see [Materials and Methods](#))

cows [26]. Intramammary challenge with *E. coli* caused massive tissue damage as observed during sample collection and in turn led to a decrease in milk production in the cows investigated. Mastitis-associated reduction of milk production and decrease of the milk protein gene expression levels following infection with *E. coli* have been reported in other studies [27–30]. In our experiment, an analogous reduction of milk protein gene expression was observed in the non-depleted (0) and the depleted samples (experimental variants A and B) in *E. coli*-challenges compared to non-challenged samples (Fig. 3).

After selective RNA depletion only 29% (25–36%) (experimental variant A) or 24% (21–30%) (experimental variant B) of the fragments from the non-challenged udder quarters were assigned to the milk protein genes (Fig. 3). The effect of the targeted RNA depletion could also be clearly observed in the challenged udder quarters; only 11% (0–24%) or 9% (0–19%) (variant A or B, respectively) of all fragments mapped to milk protein transcripts.

Altogether, the RNA-Seq analysis of the targeted depletion of milk protein gene transcripts in the RNA pool from lactating mammary tissue showed that the proportion of highly abundant milk protein gene transcripts could be successfully reduced. The RNA depletion effect was observed with both experimental

depletion conditions (A and B), independently of the challenge status of the udder tissue and resulted in a reduction of milk protein transcripts by more than 50% compared to the non-depleted sample.

Across all samples, we found that the *CSN2* transcripts were the most abundant milk protein gene transcripts (followed by *CSN1S1*, *CSN3* and *PAEP* transcripts, see Table 2) in the mammary gland of lactating cows, which confirms the results of previous transcriptome profiling reports [10].

However, focusing on individual targeted milk protein genes, we observed some variability of the RNA depletion efficacy. A nearly complete reduction of transcript levels was recorded for *CSN3* and *CSN1S2* (up to 100% of fragments) in depleted samples compared to non-depleted samples (Table 2). A 75% depletion effect was achieved for *CSN2* transcripts with a reduction from an average proportion of 16% of the counted fragments in non-depleted samples to 4% (variant A and B) after RNA depletion. Transcript depletion of *CSN1S1*, *PAEP* and *LALBA* was less comprehensive (Table 2). This indicates that not all these transcripts have been completely removed from the transcriptome after performing the selective RNA depletion procedure.

The different efficiency in the RNA depletion of the individual milk protein transcripts may have different causes, e.g. the formation of secondary or tertiary

Table 2 Percentage of fragments mapped to individual milk protein transcripts in relation to the total number of fragments in each sample as counted by featureCounts [50]

Udder quarter treatment	Animal No.	Sample treatment ^a	LALBA	CSN1S1	CSN3	CSN2	CSN1S2	PAEP
Non-challenged	Cow1	0	2.7	15.7	11.5	31.1	4.2	5.5
		A	3.1	17.9	0.6	9.6	0.3	4.5
		B	2.6	12.2	0.2	8.7	0.2	5.9
	Cow2	0	0.8	18.7	9.6	14.3	4.9	3.9
		A	0.5	18.7	0.8	3.1	0.1	2.2
		B	0.4	16.4	0.2	2.8	0.1	2.0
	Cow3	0	1.0	16.7	9.5	21.6	4.9	7.4
		A	1.0	14.3	0.7	4.2	0.1	5.2
		B	0.8	11.6	0.1	3.8	0.1	5.0
Challenged	Cow1	0	1.1	15.2	11.1	21.9	3.3	4.3
		A	1.0	12.1	0.6	5.9	0.1	3.8
		B	0.9	9.0	0.2	5.4	0.0	4.0
	Cow2	0	0.4	12.3	7.5	6.3	2.1	4.2
		A	0.2	6.2	0.2	2.2	0.0	1.8
		B	0.2	4.5	0.0	1.4	0.0	1.7
	Cow3	0	0.0	0.1	0.5	0.1	0.0	0.0
		A	0.0	0.1	0.0	0.0	0.0	0.0
		B	0.0	0.0	0.1	0.0	0.0	0.0

^aUdder tissue samples without RNA depletion (0), with RNA depletion by variant A (A) or variant B (B). Variant A and B differ in antisense oligonucleotide input for depletion (see Materials and Methods)

structures of the targeted milk protein transcript sequences, which may impede the hybridization of the selected antisense oligonucleotides to the respective sequence region. Some possible reasons, however, for the variable efficiency of the RNA depletion across the targeted milk protein genes can be excluded:

Visual inspection of targeted milk protein gene sequences revealed that there are no internal polyA stretches with a number of consecutive A nucleotides > 7 present, which could have compromised the subsequent polyA selection step. Differences in the distance between the selected oligonucleotide position and the 3' end of the target gene can also be excluded as a possible general cause for variable RNA depletion efficiency, since the *CSN2* gene was very successfully depleted, although its capture oligonucleotides are located at a distance of 284 and 439 nucleotides from the 3' end of the gene. Across all milk protein genes, the selected oligonucleotides had a medium distance of about 280 bp to the 3' end of the target gene (range from 35 to 443). However, it is conceivable that it would be more advantageous to select the capture oligonucleotides as close as possible to the 3' end of the gene to be depleted, which was not always possible due to the inherent sequence properties of individual genes (e.g., repetitive elements, known sequence variation).

Another reason for insufficient depletion of target transcripts might be genetic variants localized in the gene section covered by the oligonucleotide used for RNA depletion. However, transcript sequence analysis of the individuals included in this study revealed that SNPs in capture oligonucleotide sequences could only be detected for *CSN3*. As *CSN3* transcripts were almost completely depleted, obviously this sequence variation did not affect the depletion efficiency.

However, the RNA depletion approach may have limitations, if the animals to be analyzed show not yet detected alternative splice sites in the targeted gene. The visual inspection of read alignments using the Integrative Genomics Viewer (IGV) revealed that the lower RNA depletion efficiency observed for *CSN1S1* can be due to the existence of a high number of alternative splice variants, which were not all addressed by the capture oligonucleotides designed. The new bovine genome annotation ARS-UCD1.2 at NCBI that has been published very recently displays 25 transcript variants predicted based on now available RNA-Seq data (<https://www.ncbi.nlm.nih.gov/gene/282208>), not all of which were known at the time of primer design. Both *CSN1S1* target oligonucleotides are located in different splice sites, where an intron may be inserted, so that some highly expressed transcript variants cannot

be covered. In this case, adding additional capture oligonucleotides in the RNA depletion assay, which cover all *CSN1S1* transcript variants may be beneficial in future iterations of the depletion protocol. For the other genes with incomplete RNA depletion efficiency, *PAEP* and *LALBA*, neither alternative splice variants nor genetic variation in oligonucleotide sequences were detected.

The visual inspection of *PAEP* and *LALBA* read alignments using the IGV showed that their reads in depleted RNA samples were biased toward increased coverage at the 3' end of the respective gene. In these samples, a clear decline of read coverage is visible behind the last nucleotide of the capture oligonucleotide in 5' direction of the gene as expected. Respective non-depleted samples showed no analogous decline in read coverage, (see Fig. 4, Additional file 1). These data indicate that the observed trend of a positional bias towards an increased coverage at the 3' end of the gene may be associated with some limitations in the RNA depletion procedure. Obviously, the 3' fragments of transcripts, which carry the polyA-site and remain in the RNA pool after RNase H degradation, are still captured in the poly A+ selection step of the library preparation and finally compromise the depletion of the corresponding genes. Our data indicate that the depletion efficiency of *PAEP* and *LALBA* could possibly be improved by selecting the capture oligonucleotides as close as possible to the 3' end of the gene.

RNA depletion maintains individual mammary gland milk protein gene expression patterns in cows challenged with the pathogen *E. coli*

As already mentioned above (Fig. 3), the proportion of milk protein transcripts in the mammary gland transcriptome is lower after intramammary challenge with *E. coli* compared to the non-challenged control tissue. This can probably be explained by severe tissue damage resulting in cell destruction followed by a reduced milk protein synthesis and can also be due to an active defense against the pathogen with associated upregulation of immune defense genes. Monitoring of the expression profiles of selected transcripts coding for milk protein transcripts in each cow showed individual changes after pathogen challenge (Table 2). In all three cows, reduced expression of milk protein transcripts in non-depleted and depleted samples was evident after *E. coli* challenge, whereby cow 3 showed an almost complete lack of target gene expression after intramammary challenge. This indicates that the milk protein synthesis in the mammary gland tissue of the challenged udder quarter of this cow had almost completely stopped. Indeed, with only about 38% of the original amount of milk in the challenged udder quarter, this cow showed the largest drop in milk yield after *E. coli* challenge compared to the other two cows enrolled in this experiment. In contrast to the other two lactating cows, the pathological report of cow 3 showed granulocyte infiltration into the tissue of the challenged udder quarter. A need for RNA depletion would not have been

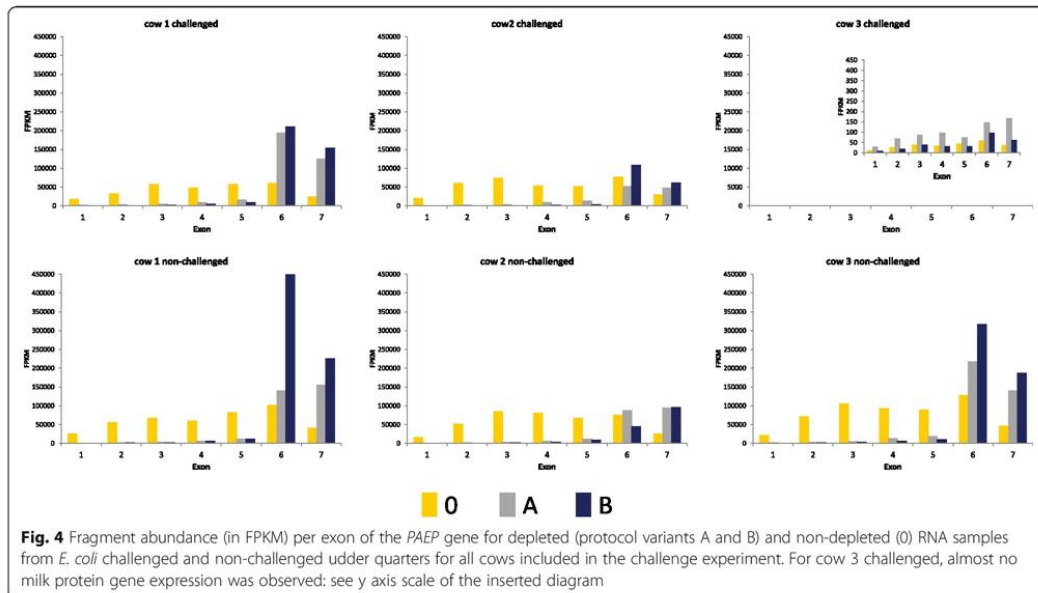


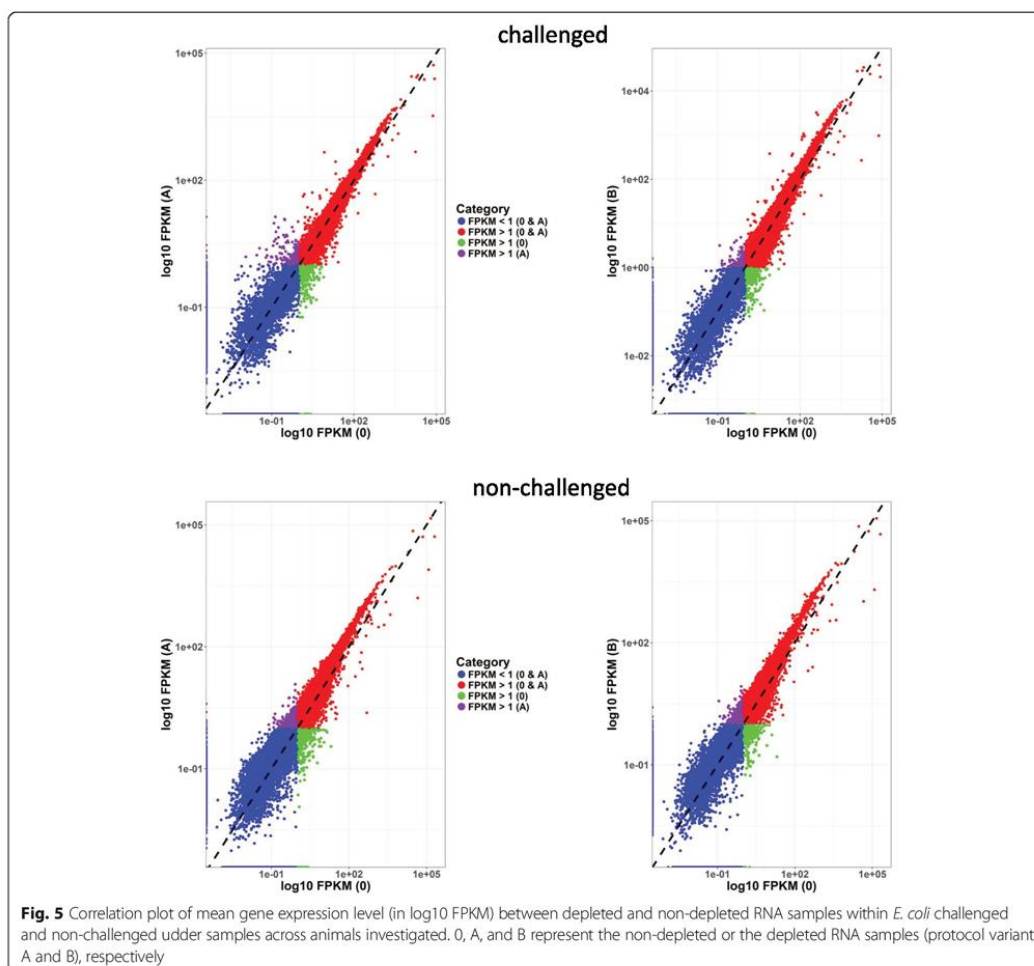
Fig. 4 Fragment abundance (in FPKM) per exon of the *PAEP* gene for depleted (protocol variants A and B) and non-depleted (O) RNA samples from *E. coli* challenged and non-challenged udder quarters for all cows included in the challenge experiment. For cow 3 challenged, almost no milk protein gene expression was observed: see y axis scale of the inserted diagram

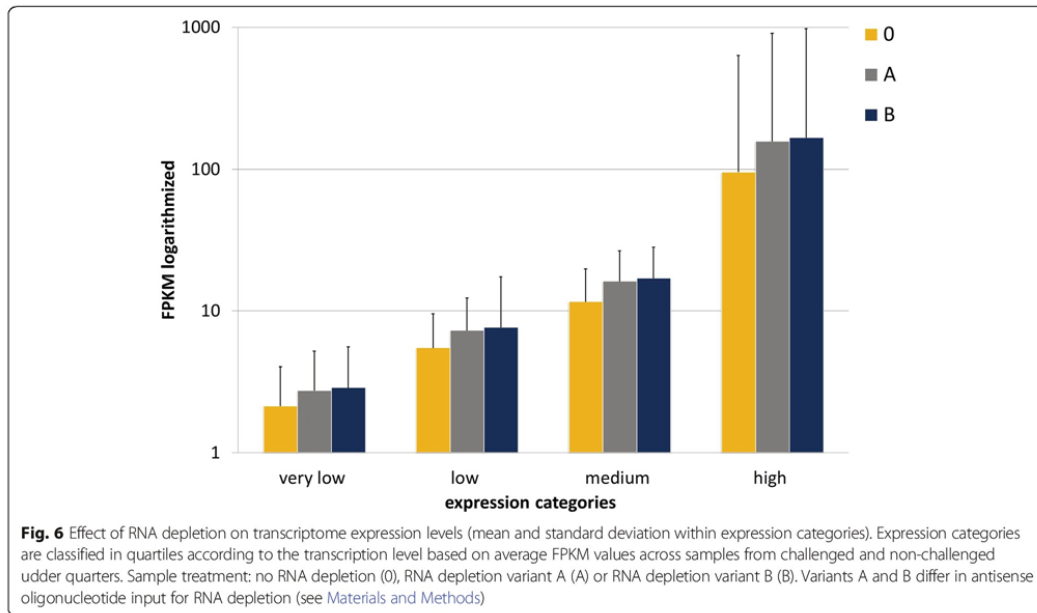
required for this sample per se, which is clearly reflected in the results obtained. However, to avoid experimentally induced bias when comparing challenged and non-challenged udder quarters in a challenge experiment, identical treatment of both samples is essential.

Effect of RNA depletion targeting highly abundant transcripts on the detection of other transcripts in the transcriptome

A correlation analysis between FPKM data for depleted and non-depleted RNAs from the same udder quarter demonstrated that the selective RNA depletion procedure did not introduce a systematic bias into gene quantification (see Fig. 5, Additional file 2). The density plots for monitoring the distribution of quantitative gene

expression revealed a shift towards higher levels for the variants A and B compared to the non-depleted samples (see Additional file 3). To evaluate the targeted RNase H-mediated RNA depletion effect on the sensitivity of RNA-Seq analysis, expression quartiles of the transcripts were formed, graded according to their transcription level. Monitoring of the RNA depletion effect on the average FPKM value of the different expression categories revealed that they shifted to higher values in consequence of RNA depletion indicating an improved sensitivity of RNA-Seq analysis (Fig. 6). The average FPKM of the “very low” transcript expression quartile has increased from 2.1 in the non-depleted samples to 2.7 (variant A) and 2.9 (variant B) after RNA depletion. In the “low expression” quartile, the average FPKM value





has risen from 5.5 to 7.2 and 7.6 in variant A and B, respectively. The FPKM values of transcripts of the “medium expression” quartile increased from 11.6 to 16.1 and 16.9, and in the “high expression” transcript quartile from 95.6 to 156.8 and 166.5, in variant A and B, respectively. The two methodological RNA depletion variants did not differ substantially (Fig. 6).

In addition, we separately explored the number of genes with an FPKM > 1 that exceeded this threshold only in depleted or only in non-depleted RNA samples. For these comparisons, we analyzed the RNA-Seq read count data based on the Ensembl 87 reference annotation. Table 3 indicates that we had a net surplus of 239 or 278 genes for the depletion groups A and B, respectively. Additional file 4 illustrates the overlap in genes with FPKM > 1 in RNA depleted and non-depleted samples.

Table 3 Number of expressed genes with FPKM > 1 and of genes with changed FPKM values after selective RNA depletion for all groups based on the Ensembl 87 reference annotation

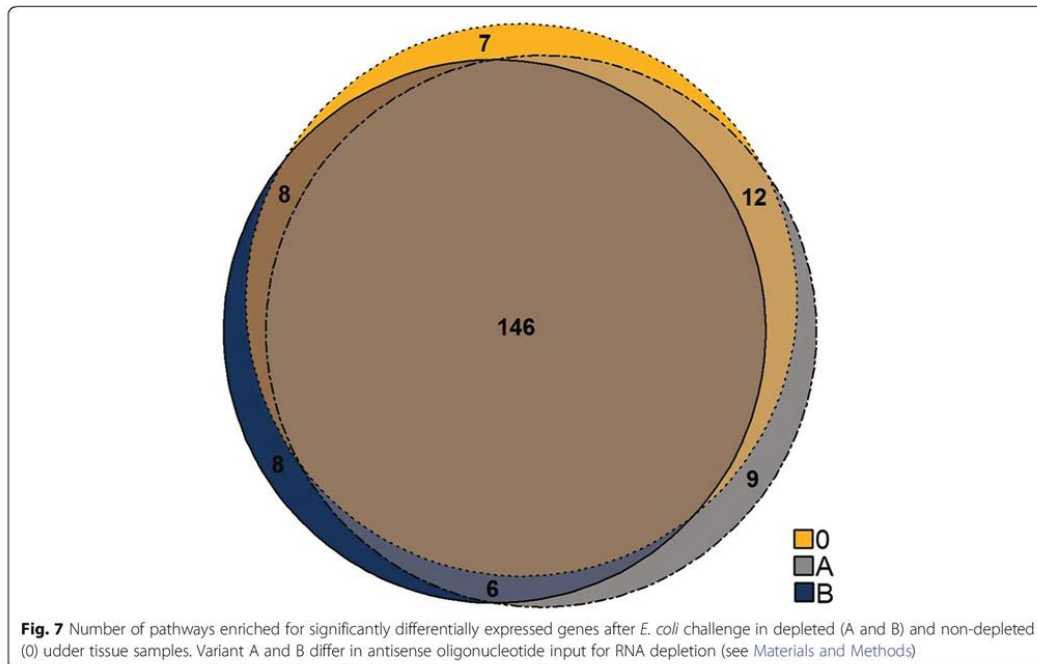
Groups ^a	All genes with FPKM > 1	Genes with FPKM > 1 only after RNA depletion	Genes with FPKM > 1 before and FPKM < 1 after RNA depletion
0	12,546	/	/
A	12,773	554	315
B	12,817	582	304

^aSamples without RNA depletion (0) and with RNA depletion according to variants A (A) and B (B). Variants A and B differ in antisense oligonucleotide input for RNA depletion (see [Materials and Methods](#))

Differential expression analysis of loci in response to intramammary challenge

To evaluate whether the depletion of highly abundant transcripts causes a bias in physiological reaction patterns in response to an *E. coli* challenge and whether the detection of significantly differentially expressed transcripts can be improved, a differential expression analysis of *E. coli* challenge vs. control was performed.

Non-depleted samples (0) showed 1079 significantly differentially expressed loci (for list of loci see Additional file 5). Targeted RNA depletion of highly abundant milk protein transcripts resulted in 1290 (variant A) and 1226 (variant B) significantly differentially expressed loci (for the list of loci see Additional file 5), which demonstrates a clear increase in number when applying the RNA depletion method. Overall, the transcriptomic determination of general physiological reaction patterns after *E. coli* udder challenge should not be compromised by the RNA depletion procedure. To monitor this, significantly differentially expressed biological pathways in challenged and non-challenged samples were analyzed based on differentially expressed loci. Ingenuity pathway analysis (for lists of significantly enriched canonical pathways see Additional file 6) showed high agreement between the groups (0, A and B) with 146 identified common biological pathways, which were consistently enriched in the transcriptome of both non-depleted (0) and depleted samples (A and B) after *E. coli* challenge (Fig. 7). The number of pathways that only occurred in one or two of the



groups was low (6–12 pathways). This indicates that the RNA depletion procedure does not result in an unintentional bias in the global biological signaling pathway patterns in response to *E. coli* challenge.

Further, we tested for potential differential expression of individual genes well described in literature to be affected by *E. coli* challenge in the mammary gland [31]. We had a specific focus on immune response processes and looked particularly at genes encoding cytokines as well as chemokine and inflammatory relevant receptors. The comparison of the transcription profiles in response to the *E. coli* challenge between the methodological groups (0, A and B) showed that RNA depletion improved the sensitivity to detect such well-known effects on gene expression. We observed that significant differential expression of some cytokine encoding genes in response to *E. coli* challenge was only detectable in the depleted samples A and B in contrast to the non-depleted samples.

This included the master regulator of acute phase response *IL6* gene [32, 33] (Table 4) and the *CCL16* gene. The *CCL20* gene was found to be significantly differentially expressed only in samples processed with RNA depletion variant B. These cytokines and chemokines are known to be regulated after challenge with *E. coli* [34–36]. For relevant transmembrane receptors, we observed a challenge-related, significantly different expression only in depleted samples. This applied to *TREM1* (triggering

receptor expressed on myeloid cells 1) gene, a critical regulator of diverse cellular functions including amplification of inflammation [37], and the *CCR2* and *CCR7* genes, both encoding G-protein coupled receptors, which are essential for leukocyte recruitment [36]. A targeted differential RT-qPCR analysis for the *IL6* and *CCL20* genes between pathogen-challenged and non-challenged tissue samples confirmed a higher sensitivity for detecting a response to pathogens in depletion variant B (Fig. 8).

The highly abundant milk protein transcripts obviously seem to impede the detection of these exemplarily selected, lowly expressed transcripts in the analysis of transcriptomes using RNA-Seq. Thus, the depletion of the milk protein transcripts from the total RNA pool prior to RNA-Seq offers a higher sensitivity for detecting relevant biological signaling processes in global transcriptome analysis of mammary gland tissue from lactating cows, and thus enables or improves the detection of genes expressed at a lower level.

Detection of unknown genes

By improving the sensitivity of a RNA-Seq analysis due to selective depletion of highly abundant transcripts, it is expected that also an improvement in detection of yet unknown transcripts will be achieved. After generating transcriptome annotations separately for the non-depleted and depleted experimental group (0, A, B, without

Table 4 Differential expression obtained from RNAseq data for selected genes in non-challenged udder samples compared to *E. coli* challenged samples

Gene	q-value ¹		
	O ²	A ²	B ²
<i>IL6</i>	0.092	0.019	0.002
<i>CCL20</i>	0.662	0.289	0.002
<i>CCL16</i>	0.057	0.019	0.004
<i>TREM1</i>	0.096	0.036	0.002
<i>CCR2</i>	0.189	0.200	0.022
<i>CCR7</i>	0.124	0.051	0.024

¹ Statistical significance (q value) for differential expression between infected and non-infected samples

² Udder tissue samples without RNA depletion (O), with RNA depletion according to variants A (A) and B (B). Variants A and B differ in antisense oligonucleotide input for depletion (see Materials and Methods)

considering the challenge status), the total number of expressed loci found in the transcriptomes of each group was determined. Non-depleted samples showed a lower number of expressed loci than samples depleted with variant A or B (Table 5). Even after applying an expression threshold value of FPKM > 1, a clearly higher number of loci were detected in depleted samples than in non-depleted samples. A similar result was obtained when analyzing for yet unknown loci. For RNA-depleted samples, a higher number of these loci yet unannotated in the *Bos taurus* Ensembl annotation were identified compared to non-depleted samples, also at a threshold value of FPKM > 1. The application of our RNA-depletion approach thus provides a generally higher number of discovered loci and facilitates the detection and analysis of new loci.

Conclusions

The results of our study suggest that the sensitivity of the RNA-Seq analysis is improved when removing highly abundant milk protein gene transcripts from the total RNA pool of the mammary gland from lactating cows prior to RNA-Seq library preparation (negative selection). The implementation of a selective RNase H-mediated RNA depletion of milk protein gene transcripts from the transcriptome of the mammary gland tissue of lactating cows will help to achieve a more comprehensive transcript catalogue of the mammary gland transcriptome, which better reflects its complexity. At the same time, our data confirm that targeted milk protein transcript depletion does not introduce a bias in the outcome of transcriptome analyses within mammary gland pathogen challenge experiments.

Depletion variant B is superior to A for a number of parameters: stronger depletion of milk protein transcripts (Fig. 3, Table 2), higher number of genes (from the reference annotation Ensembl 87) with FPKM > 1 not found in non-depleted samples (Table 3), higher

number of differentially expressed genes found in *E. coli* challenge (Additional file 5) and small, but consistently higher FPKM values for gene expression quantiles (from the reference annotation Ensembl 87, Fig. 6). This suggests that a higher and equimolar concentration of the antisense oligonucleotides in RNA depletion assay should be applied in follow-up studies.

Methods

Animals and samples

Twelve Holstein Friesian cows were challenged during their first lactation with *E. coli* for 24 h [38]. The animals were purchased from ordinary commercial dairy farms in the northeast region of Germany. One udder quarter of each cow was challenged with 500 colony forming units of *E. coli* 1303. The intramammary challenge was performed at day 36 ± 3.4 after first parturition, 24 h before dissection of the cow. One of the remaining untreated udder quarters was used as control. The cows were separately housed in a loose-stall barn and milked twice daily. Environmental conditions and feeding during the calving period and the challenge interval were identical for all cows enrolled in the study as previously described [38]. Of these challenged 12 animals, three lactating cows without clinical signs of diseases at the start of the intramammary challenge, but with different degrees of clinical response to pathogen challenge were used to validate the RNase H-mediated RNA depletion method targeting milk protein genes highly expressed in the mammary gland. Two of the cows included in this experiment (cow 1 and 3) are offspring from the same sire, but the maternal ancestry is different for all cows. Mammary gland parenchymal tissue of each quarter from these three cows was collected immediately after killing and dissection, frozen in liquid nitrogen and stored at -80 °C.

RNA preparation

For each of the three cows, frozen samples (50 mg) of parenchymal tissue from the challenged and from a control (non-challenged) udder quarter were pulverized in liquid nitrogen, and total RNA was extracted with TRIzol reagent (Invitrogen, Darmstadt, Germany) followed by an on-column-purification using the NucleoSpin RNA II kit (Macherey & Nagel, Düren, Germany) with modifications of the DNase digestion step according to Weikard et al. [39]. After testing the total RNA preparation for genomic DNA presence by PCR [40], the DNase-treatment step was repeated when necessary. The RNA concentration was measured with a Qubit Fluorometer (Invitrogen, Germany). For RNA quality control, RNA integrity was determined using the 2100 Bioanalyzer Instrument (Agilent Technologies, Germany).

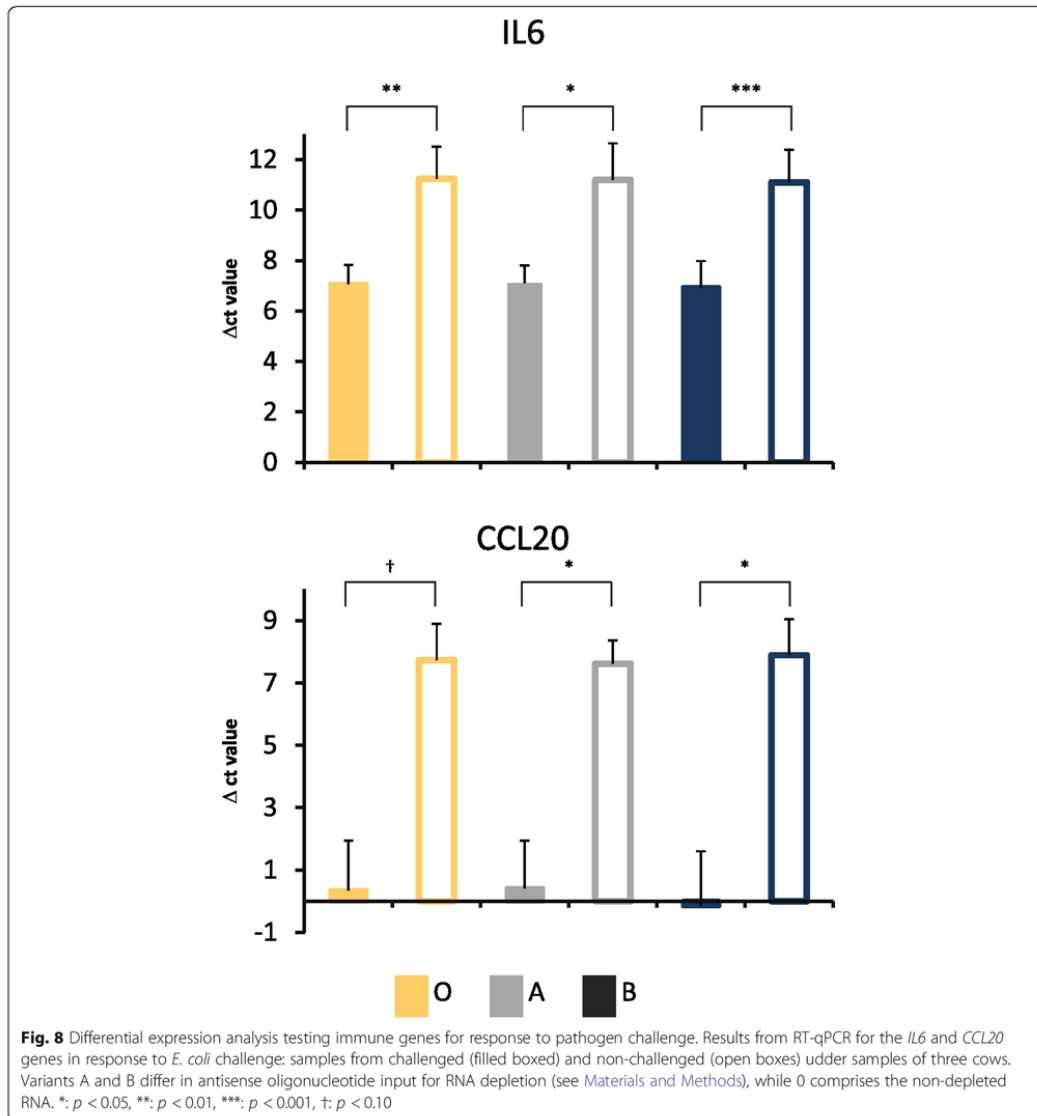


Table 5 Number of all expressed loci and the subset of unknown expressed loci based on a group-specific transcriptome annotation

Groups ^a	All loci	All loci with FPKM > 1	Unknown loci	Unknown loci with FPKM > 1
O	32,434	16,168	4411	1915
A	43,331	19,396	11,023	2997
B	35,510	17,779	6343	2509

^aSamples without RNA depletion (O) and with RNA depletion according to variants A (A) and B (B). Variants A and B differ in antisense oligonucleotide input for depletion (see [Materials and Methods](#))

Depletion of highly abundant transcripts

The depletion procedure of highly abundant transcripts originating from the casein gene cluster (*CSN1S1*, *CSN1S2*, *CSN2* and *CSN3*), α lactalbumin gene (*LALBA*) and progesterone associated endometrial protein gene (*PAEP*) in lactating mammary gland tissue is based on hybridization of antisense oligonucleotides specific for these target transcripts to total RNA prior to library preparation used for RNA-sequencing (RNA-Seq). The targeted RNA depletion procedure is presented in Fig. 1. Two antisense oligonucleotides with a melting temperature $> 65^\circ\text{C}$ were derived of each target gene reference sequence, *Bos taurus* reference genome assembly UMD3.1 (see Table 6) using the OLIGO Primer Analysis Software (MedProbe, Oslo, Norway). For the selection of antisense oligonucleotides that are specific for the targeted milk protein genes, care was taken that they are located as far as possible in exons located close to the 3' end of the gene to be depleted and outside of repetitive and low complexity sequence regions as well as known genetic variants and alternative splice sites of the targeted genes. To avoid off-target amplification the specificity of the selected oligonucleotides (Table 6) was checked by BLAST search against the *Bos taurus* reference transcriptome and genome assembly (UMD3.1, annotation release 105 [41]) using the Primer-BLAST tool [42].

The effect of concentrations of the antisense oligonucleotides in the hybridization assay was tested in two different experimental conditions (Table 7). In the first variant (A), the final oligonucleotide concentration was adjusted according to the expression level of the respective target gene in lactating mammary gland known from other studies [10]. In the second variant (B), a constant final equimolar concentration of 25 μM was applied for each oligonucleotide in the hybridization assay (Table 7).

Duplicates of 1 μg of total RNA from each sample were denatured by incubation for 2 min at 95°C in

hybridization buffer (10 mM Tris-HCl, pH 7.6, 20 mM KCl) and hybridized for 5 min at 65°C with antisense oligonucleotides pooled in a total volume of 10 μl . The reactions were carried out in a thermocycler and completed with cooling down to 4°C .

Immediately after hybridization of the RNA samples with the antisense oligonucleotides, a selective digestion of the RNA strand of the RNA-oligonucleotide hybrids was carried out with RNase H (#AM2292, Ambion). The RNase H reaction mixture consisting of 2 U RNase H and 20 U SUPERase-In (inhibiting RNase A, B, C, 1 and T1, Ambion, #AM2694) in RNase H-buffer (20 mM Tris-HCl, pH 7.6, 4 mM MgCl_2 , 0.02 mM DTT) was prepared on ice. The hybridized RNA-oligonucleotide hybrids were added to 10 μl of RNase H reaction mixture, mixed, incubated at 37°C for 10 min and cooled down to 4°C . To stop the reaction, 1 μl 0.5 M EDTA was added immediately. After shortly spinning down the reaction mixtures the respective duplicates of the samples were pooled, and finally the pooled samples were purified with the RNeasy MinElute Cleanup kit (Qiagen). The RNA quality was controlled with the Agilent Bioanalyzer 2100 and the RNA quantification was carried out using the Qubit Fluorometer.

Monitoring of the RNA depletion effect by RT-qPCR

Depleted and non-depleted RNA samples were reverse transcribed to cDNA using the SuperScript First-Strand Synthesis System III (Invitrogen, Thermo Fisher Scientific) according to the manufacturer's instructions. RT-qPCR analysis was performed as described (Weikard et al. 2012) on a LightCycler qPCR platform (Roche). Milk protein gene copy numbers were normalized against those of *RPS15A* used as a reference gene. Primers for transcript quantification via qPCR of each target gene were designed based on the respective reference gene sequence, *Bos taurus* reference genome assembly UMD3.1 (see Table 8).

Table 6 Antisense oligonucleotide sequences designed for targeting highly abundant transcripts in the mammary gland

Gene name	Accession number	Oligonucleotide name	Oligonucleotide sequence (5'-3')
<i>PAEP</i>	NM_173929.3	LGB_R5	GGGCTCACCTAGATGTGGCACTGCTC
		LGB_R6	GCTCAGCACTGTTCTCCATGCAGAAG
<i>LALBA</i>	NM_174378.2	LALBA_R6	GTATCAGTAAGATCATCTCCAGGAAC
		LALBA_R7	CAGAACAGAGTGCTTTATGGGCCAAC
<i>CSN1S2</i>	NM_174528.2	CSN1S2_R3	ATAACCAGGTAGAAGCAGTTAATCCAG
		CSN1S2_R4	ATGCTGGTTGTATGAAGTAAAGTGGTAG
<i>CSN1S1</i>	NM_181029.2	CSN1S1_R7	TCAGAATTCACCTTGACTCTCACCACAG
		CSN1S1_R8	TAGGATTAGGGATGTCAGAGAATGATGG
<i>CSN3</i>	NM_174294.2	CSN3_R6	GTAAGAGGAGACGAGGAAGGAGCCAG
		CSN3_R7	ACTGTGTTGATCTCAGGTGGGCTCTC
<i>CSN2</i>	XM_010806178.1	CSN2_R4	TCCAGTCGAGTCAATTCAAAAGTGAG
		CSN2_R5	CTTCTGGGGAACAGGCAGGACTTTG

Table 7 Final concentration of the antisense oligonucleotides in the RNA depletion assays

Oligonucleotide pair	Concentration [μ M]	
	Variant A	Variant B
LGB_R5, LGB_R6	16.66	25.0
LALBA_R6, LALBA_R7	16.66	25.0
CSN1S2_R3, CSN1S2_R4	8.33	25.0
CSN1S1_R7, CSN1S1_R8	16.66	25.0
CSN3_R6, CSN3_R7	9.0	25.0
CSN2_R4, CSN2_R5	25.0	25.0

Primer pair specificity was checked by BLAST search against the *Bos taurus* reference transcriptome and genome assembly (UMD3.1, annotation release 105 [41]) using the Primer-BLAST tool [42]. Sequences for gene-specific PCR primers are provided in Table 8.

Library preparation and RNA sequencing (RNA-Seq)

The effect of the RNA depletion on the transcript composition of the respective samples was monitored by comparative transcriptome analysis of depleted (variant A and B) and non-depleted RNA samples (0), from the challenged and an non-challenged control udder quarter using RNA-Seq. Stranded indexed poly(A+) selected libraries were prepared from 250 ng depleted or non-depleted

RNA of mammary gland samples using the TruSeq RNA Library Preparation Kit v2 set A (Illumina). The standard procedures for Illumina's mRNA-Seq were applied, and the libraries were subjected to paired-end (2×100 bp) mRNA sequencing in two lanes on the HiSeq 2500 Sequencing System (Illumina) [43].

Read mapping

After removal of adapters with Cutadapt (version 1.13) [44], quality control with FastQC [45] and read quality trimming with QualityTrim (version 1.6.0) [46], read mapping was performed with HISAT2 [47] against the bovine reference genome (UMD3.1 [48], Ensembl-Annotation release 87 [49]).

Evaluation of the RNA depletion success

To assess the success of the targeted RNA depletion of the milk protein gene transcripts, the proportion of these transcripts in the transcriptomes of depleted (variant A and B) and non-depleted control samples of challenged and control udder tissue was determined. FeatureCounts [50] was used to calculate sequencing fragments assigned to the loci present in the *Bos taurus* genome annotation (Ensembl-Annotation release 87). First, all sequencing fragments of a sample were summed up (=100%) using R, version 3.4.3 [51]. Afterwards, only those fragments, which

Table 8 Sequences of primers used in RT-qPCR

Gene name	Gene accession number	Oligonucleotide name	Oligonucleotide sequence (5'-3')
PAEP	NM_173929.3	LGB_F1	CAAGATCCCTGCGGTGTTCAAG
		LGB_R1	ACTGTTCTCCATGCAGAAGAGC
LALBA	NM_174378.2	LALBA_F2	CAGTTTGCTGAATGGGTCT
		LALBA_R3	GATCATCATCCAGGAAGTGTGTC
CSN1S2	NM_174528.2	CSN1S2_F2	CACCACTGAGGAAAATTCAAAG
		CSN1S2_R1	GCTGATAACAGCTTTGAGATAC
CSN1S1	NM_181029.2	CSN1S1_F4	GTGCTGAGGAACGACTTCACAG
		CSN1S1_R4A	CAGTGTATTGTGCTAGTGG
CSN3	NM_174294.2	CSN3_F5	GCCCACCTGAGATCAACACAG
		CSN3_R5	AGGAGACGAGGAAGGAGCCAG
CSN2	XM_010806178.1	CSN2_F1	TGAGGAACAGCAGCAACAG
		CSN2_R1	ACAGGGGTTTGAGTAAGAGG
RPS15A	NM_001037443.1	RPS15A_F1	CCGTGCTCCAAAGTCATCGTC
		RPS15A_R1	TGAAACCAAAGTACGGGATG
CCL20	NM_174263.2	CCL20_F3	CAGCAAGTCAGAAGCAAGCAA
		CCL20_R1	CCCACCTCTCTTTGGATCTGC
IL6	NM_173923.2	IL6_F1	GGAGGAAAAGACGGATGCT
		IL6_R1	GGTCAGTGTGCTGGTGA
PPP1CC	NM_174581.2	PPP1CC_F1	TGCAGCATCATCCAACG
		PPP1CC_R1	CGGAAACCACTGACTCA

were assigned to the target genes (*CSN1S1*, *CSN1S2*, *CSN2*, *CSN3*, *LALBA* and *PAEP*) were counted, and the respective fraction across all target genes was calculated for each sample. Finally, averages of all three animals for control and depleted samples were calculated from these values and compared. In addition to average values across all genes in depleted and non-depleted, challenged and non-challenged samples, fragments for single milk protein genes of each udder quarter were analyzed and compared for a more detailed evaluation.

Analysis of expression categories

To determine if/how the general expression patterns of genes of the mammary transcriptome were affected by the RNA-depletion procedure, the transcripts present in the mammary transcriptome of lactating udder tissue were grouped into expression categories based on their transcript expression levels.

Fragments per kilobase per million fragments mapped (FPKM) were calculated for all loci based on the fragment counts determined by featureCounts. To define the limits of the expressions categories, an average FPKM for each locus across all non-depleted samples (challenged and non-challenged) was calculated. All loci with an average FPKM value lower than 1.0 and the six milk protein genes targeted by RNA depletion were removed from the dataset.

From this final dataset, 25, 50 and 75%- quartiles were calculated. The limits of expression categories were defined accordingly. The expression quartiles comprised transcripts with FPKM values ranging from 1 to 3.55 FPKM ("very low"), 3.55 to 7.75 FPKM ("low"), 7.75 to 17.19 FPKM ("medium") and 17.19 to 21,068.1 FPKM ("high").

In the next step, average FPKM values of all transcripts of each defined expression quartile were calculated. This was performed for the non-depleted as well as the depleted samples (variants A and B) separately, but with challenged and non-challenged samples together.

Analysis of depletion effect on gene expression level by read counts from RNA-Seq

Read counts per gene as obtained by featureCounts from the Subread package [50] based on the reference Ensembl 87 annotation were further screened for a potential bias introduced into gene quantification by correlation analysis between depleted and non-depleted RNA for each udder quarter and also within challenged and non-challenged groups. Furthermore, we also performed an exon-wise expression analysis of all milk protein genes targeted by depletion and inspected the alignment files by Integrated Genomics Viewer, IGV [52].

Evaluation of the RNA depletion effects on the *E. coli* challenge response

Investigation of differential gene expression in response to *E. coli* challenge and whether this is affected by the targeted RNA depletion of highly abundant transcripts was performed using Cuffdiff (v.2.2.1) [53] with default parameter settings based on the *Bos taurus* UMD 3.1 Ensembl v87 genome annotation. The differential gene expression of challenged versus non-challenged udder tissue samples was compared in both experimental RNA depletion variants (A and B) and the control group separately (0), respectively. Thereafter, using R-scripts [51] the total number of significantly differentially expressed loci ($q < 0.05$) in response to pathogen challenge exceeding a threshold of FPKM = 1 were counted and compared between technical groups (control, variant A, variant B).

The Cuffdiff output of each group (0, A and B) was also used for biologically functional pathway analyses performed with the Ingenuity Pathway Analysis software (IPA) [54]. IPA analysis was performed on lists of loci identified as significantly differentially expressed ($q < 0.05$). Results of this analysis are presented in a venn diagram, created with the eulerr package [55] in R [51].

Targeted RT-qPCR for genes with different expression in response to *E. coli* challenge

Depleted and non-depleted RNA samples were reverse transcribed to cDNA and RT-qPCR analysis was performed essentially as described above. Primer sequences of transcripts subjected to quantification are provided in Table 8. Gene expression levels (cycle quantification, Cq values), were normalized against that of the *PPP1CC* gene serving as reference gene. Differential expression between challenged and non-challenged samples was analyzed by t-tests within non-depleted and RNA-depleted samples.

Detection of unknown genes

To evaluate the RNA depletion effect on the capability to detect yet unknown loci, we used StringTie [56] to perform a reference-based (UMD3.1 [46], Ensembl annotation release 87 [47]) transcriptome assembly for each group (0, A, B) separately. The mapped reads of each dataset of each animal were assembled and the individual gtf files subsequently merged according to the groups, resulting in three group-specific transcriptome annotations.

Based on these annotations the total number of loci expressed in each group was determined with featureCounts. To this end, an average FPKM value for each locus was calculated. To improve the specificity of the data and remove potential background noise those loci with an FPKM < 1 were removed from the dataset.

In addition, the group-specific assemblies were evaluated with regard to unknown/new loci with GffCompare [57]. The detected transcripts of the output were computed into loci and the number of those loci with the class code “u” (unknown) and consisting of more than one exon were counted and compared between the groups.

Additional files

Additional file 1: IGV screen shot of the read distribution across the *PAEP* gene for the *E. coli* challenged udder sample from cow 2. O, A and B represent the non-depleted or the depleted RNA samples (protocol variant A and B), respectively. Exon 6 is zoomed in for better demonstration of the position of the antisense capture oligonucleotide (see red line). (PDF 235 kb)

Additional file 2: Correlation plot of gene expression level (in log₁₀ FPKM) between depleted and non-depleted RNA samples within *E. coli* challenged and non-challenged udder samples of each cow investigated. O, A, and B represent the non-depleted or the depleted RNA samples (protocol variant A and B), respectively. (PDF 1551 kb)

Additional file 3: Distribution of gene expression level (FPKM) for the depleted and non-depleted RNA samples from non-challenged udder samples. O, A, and B represent the non-depleted or the depleted RNA samples (protocol variant A and B), respectively. (PDF 283 kb)

Additional file 4: Number and overlap of expressed genes with FPKM > 1 in depleted (protocol variants A and B) and non-depleted (O) RNA samples. (JPG 645 kb)

Additional file 5: Lists of differentially expressed genes comparing pathogen-challenged and non-challenged udder quarters for non-depleted (O) and depleted RNA samples (protocol variants A and B) at $q < 0.05$ used for upload into IPA pathway analysis. (XLSX 70 kb)

Additional file 6: Lists of IPA canonical pathways significantly enriched for differentially expressed genes in non-depleted (O) and depleted RNA samples (protocol variants A and B) from pathogen-challenged and non-challenged udder quarters. Pathways are classified for being observed across all protocols (O, A or B) or shared between two or all three protocols. (PDF 214 kb)

Abbreviations

CCL16: C-C motif chemokine ligand 16; CCL20: C-C motif chemokine ligand 20; CCR2: C-C motif chemokine receptor 2; CCR7: C-C motif chemokine receptor 7; CSN1S1: Casein α S1; CSN1S2: Casein α S2; CSN2: Casein β ; CSN3: Casein κ ; *E. coli*: *Escherichia coli*; FPKM: Fragments per kilobase per million fragments mapped; IGV: Integrative Genomics Viewer; IL6: Interleukin 6; IPA: Ingenuity pathway analysis; LALBA: α lactalbumin; PAEP: Progesterone-associated endometrial protein; PPP1CC: Protein phosphatase 1 catalytic subunit γ ; RNA-Seq: RNA sequencing; RPS15A: Ribosomal protein S15a; rRNA: ribosomal RNA; RT-qPCR: Reverse transcriptase-quantitative polymerase chain reaction; TREM1: Triggering receptor expressed on myeloid cells 1

Acknowledgements

The excellent technical assistance of Simone Wöhl, Bärbel Pletz, Sebastian Gaedcke, Manuela Lötze and Emma Schröder is thankfully acknowledged. We thank the Association for Bioeconomy Research (FBF) for continuous support. The project would not have been possible without the skilled and dedicated support of the staff in animal experimental facilities at the University of Veterinary Medicine, Hannover.

Funding

This work is supported by Funds of the German Government's Special Purpose Fund (Grant number 740718) held at Landwirtschaftliche Rentenbank. The funding body had no role in the design of the study, collection, analysis, and interpretation of data nor in writing the manuscript. Furthermore, the funders had no role in the decision to publish the manuscript. The publication of this article was funded by the Open Access Fund of the Leibniz Institute for Farm Animal Biology (FBN).

Availability of data and materials

RNA-Seq datasets are submitted to the ENA repository (<https://www.ebi.ac.uk/ena>, Project number PRJEB30263, accession numbers ERR2985659, ERR2987284-ERR2987300) at EMBL-EBI.

Authors' contributions

This methodological study is part of a comprehensive, collaborative study on bovine mastitis (ChronMas project) conceived and designed by HZ, WP, MH, HJS, SE and CK. MM and AH performed the main animal experiments, coordinated the clinical sample collection and took care of the animals. CK and RW conceived and designed the methodological RNA depletion study. RW developed the RNA depletion procedure. JB performed the experiments (RNA depletion, RNA sequencing) and data analysis. UT, FH, AH and CK participated in data analysis. JG, CK, HJS, HZ and RW contributed to the data interpretation, and CK supervised the study. JB, RW and CK wrote the manuscript. MM, WP, HJS and SE reviewed and edited the manuscript. All authors read and approved the final manuscript.

Ethics approval

The experiments were performed under authorization of the Lower Saxony Federal State Office for Consumer Protection and Food Safety, Germany (reference number 33.12-42502-04-15/2024). Furthermore, this study was approved by the Ethics Committee of the University of Veterinary Medicine Hannover, foundation, Germany. All ethical evaluations were performed as required by the German Animal Welfare Act (Tierschutzgesetz, <https://www.gesetze-im-internet.de/tierschg/BjNR012770972.html>). All procedures were in compliance with federal and state ethics guidelines of animal care and use and animal handling guidelines of Germany's Animal Welfare Act were followed.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interest.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Leibniz Institute for Farm Animal Biology (FBN), Institute of Genome Biology, 18196 Dummerstorf, Germany. ²Clinic for Ruminants with Ambulatory and Herd Health Services, Centre for Clinical Veterinary Medicine, Ludwig-Maximilians-University Munich, 85764 Oberschleissheim, Germany. ³Clinic for Cattle, University of Veterinary Medicine, Foundation, 30173 Hannover, Germany. ⁴Immunology Unit, University of Veterinary Medicine, Foundation, 30559 Hannover, Germany. ⁵Institute for Microbiology, Technical University Braunschweig, 38106 Braunschweig, Germany. ⁶Microbial Proteomics, Helmholtz Centre for Infection Research, 38124 Braunschweig, Germany. ⁷Agricultural and Environmental Faculty, University Rostock, 18059 Rostock, Germany.

Received: 14 December 2018 Accepted: 7 May 2019

Published online: 22 May 2019

References

1. Zadoks RN, Middleton JR, McDougall S, Katholm J, Schukken YH. Molecular epidemiology of mastitis pathogens of dairy cattle and comparative relevance to humans. *J Mammary Gland Biol Neoplasia*. 2011;16(4):357–72.
2. Fávero S, Portillo FVR, Oliveira ACR, Langoni H, Pantoja JCF. Factors associated with mastitis epidemiologic indexes, animal hygiene, and bulk milk bacterial concentrations in dairy herds housed on compost bedding. *Livest Sci*. 2015;181:220–30.
3. Rollin E, Dhuyvetter KC, Overton MW. The cost of clinical mastitis in the first 30 days of lactation: an economic modeling tool. *Prev Vet Med*. 2015;122(3):257–64.
4. Seegers H, Fourichon C, Beaudeau F. Production effects related to mastitis and mastitis economics in dairy cattle herds. *Vet Res*. 2003;34(5):475–91.

5. Piepers S, De Vliegher S, de Kruif A, Opsomer G, Barkema HW. Impact of intramammary infections in dairy heifers on future udder health, milk production, and culling. *Vet Microbiol.* 2009;134(1–2):113–20.
6. Huijps K, Lam TJ, Hogeveen H. Costs of mastitis: facts and perception. *J Dairy Res.* 2008;75(1):113–20.
7. van Soest FJS, Santman-Berends I, Lam T, Hogeveen H. Failure and preventive costs of mastitis on Dutch dairy farms. *J Dairy Sci.* 2016;99(10):8365–74.
8. Nyman AK, Emanuelson U, Gustafsson AH, Persson Waller K. Management practices associated with udder health of first-parity dairy cows in early lactation. *Prev Vet Med.* 2009;88(2):138–49.
9. Moosavi M, Mirzaei A, Ghavami M, Tamadon A. Relationship between season, lactation number and incidence of clinical mastitis in different stages of lactation in a Holstein dairy farm. *Vet Res.* 2014;5(1):13–9.
10. Canovas A, Rincon G, Bevilacqua C, Islas-Trejo A, Brenaut P, Hovey RC, Boutinaud M, Morgenthaler C, VanKlompenberg MK, Martin P, et al. Comparison of five different RNA sources to examine the lactating bovine mammary gland transcriptome using RNA-sequencing. *Sci Rep.* 2014;4:5297.
11. Wickramasinghe S, Rincon G, Islas-Trejo A, Medrano JF. Transcriptional profiling of bovine milk using RNA sequencing. *BMC Genomics.* 2012;13:45.
12. Zaheer R, Noyes N, Ortega Polo R, Cook SR, Marinier E, Van Domselaar G, Belk KE, Morley PS, McAllister TA. Impact of sequencing depth on the characterization of the microbiome and resistome. *Sci Rep.* 2018;8(1):5890.
13. Lawless N, Foroushani AB, McCabe MS, O'Farrelly C, Lynn DJ. Next generation sequencing reveals the expression of a unique miRNA profile in response to a gram-positive bacterial infection. *PLoS One.* 2013;8(3):e57543.
14. Lewandowska-Sabat AM, Hansen SF, Solberg TR, Osteras O, Heringstad B, Boysen P, Olsaker I. MicroRNA expression profiles of bovine monocyte-derived macrophages infected in vitro with two strains of *Streptococcus agalactiae*. *BMC Genomics.* 2018;19(1):241.
15. Ma Q, Li L, Tang Y, Fu Q, Liu S, Hu S, Qiao J, Chen C, Ni W. Analyses of long non-coding RNAs and mRNA profiling through RNA sequencing of MDBK cells at different stages of bovine viral diarrhoea virus infection. *Res Vet Sci.* 2017;115:508–16.
16. Herbert ZT, Kershner JP, Butty VL, Thimmapuram J, Choudhary S, Alekseyev YO, Fan J, Podnar JW, Wilcox E, Gipson J, et al. Cross-site comparison of ribosomal depletion kits for Illumina RNAseq library construction. *BMC Genomics.* 2018;19(1):199.
17. Adiconis X, Borges-Rivera D, Satija R, DeLuca DS, Busby MA, Berlin AM, Sivachenko A, Thompson DA, Wysoker A, Fennell T, et al. Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nat Methods.* 2013;10(7):623–9.
18. Krjutskov K, Koel M, Roost AM, Katayama S, Einarsdottir E, Jouhilahti EM, Soderhall C, Jaakma U, Plaas M, Vesterlund L, et al. Globin mRNA reduction for whole-blood transcriptome sequencing. *Sci Rep.* 2016;6:31584.
19. Bayatti N, Cooper-Knock J, Bury JJ, Wyles M, Heath PR, Kirby J, Shaw PJ. Comparison of blood RNA extraction methods used for gene expression profiling in amyotrophic lateral sclerosis. *PLoS One.* 2014;9(1):e87508.
20. Mastrokolas A, den Dunnen JT, van Ommen GB, 't hoen PA, Van Rooijen WM. Increased sensitivity of next generation sequencing-based expression profiling after globin reduction in human blood RNA. *BMC Genomics.* 2012;13(1):28.
21. Sultan M, Amstislavskiy V, Risch T, Schuette M, Dökel S, Ralsler M, Balzereit D, Lehrach H, Yaspo M-L. Influence of RNA extraction methods and library selection schemes on RNA-seq data. *BMC Genomics.* 2014;15(1):675.
22. Paten AM, Duncan EJ, Pain SJ, Peterson SW, Kenyon PR, Blair HT, Dearden PK. Functional development of the adult ovine mammary gland—insights from gene expression profiling. *BMC Genomics.* 2015;16:748.
23. Ibeagha-Awemu EM, Li R, Ammah AA, Dudemaine PL, Bissonnette N, Benchaar C, Zhao X. Transcriptome adaptation of the bovine mammary gland to diets rich in unsaturated fatty acids shows greater impact of linseed oil over safflower oil on gene expression and metabolic pathways. *BMC Genomics.* 2016;17:104.
24. Burvenich C, Van Merris V, Mehrzad J, Diez-Fraile A, Duchateau L. Severity of *E. coli* mastitis is mainly determined by cow factors. *Vet Res.* 2003;34(5):521–64.
25. Hagiwara S, Mori K, Okada H, Oikawa S, Nagahata H. Acute *Escherichia coli* mastitis in dairy cattle: diagnostic parameters associated with poor prognosis. *J Vet Med Sci.* 2014;76(11):1431–6.
26. Oliver SP, Calvinho LF. Influence of inflammation on mammary gland metabolism and Milk composition. *J Anim Sci.* 1995;73(suppl.2).
27. Mitterhuemer S, Petzl W, Krebs S, Mehne D, Klanner A, Wolf E, Zerbe H, Blum H. *Escherichia coli* infection induces distinct local and systemic transcriptome responses in the mammary gland. *BMC Genomics.* 2010;11(1):138.
28. Younis S, Javed Q, Blumenberg M. Meta-analysis of transcriptional responses to mastitis-causing *Escherichia coli*. *PLoS One.* 2016;11(3):e0148562.
29. Gröhn YT, Wilson DJ, González RN, Hertl JA, Schulte H, Bernet G, Schukken YH. Effect of pathogen-specific clinical mastitis on milk yield in dairy cows. *J Dairy Sci.* 2004;87:3358–74.
30. Heikkilä AM, Liski E, Pyörälä S, Taponen S. Pathogen-specific production losses in bovine mastitis. *J Dairy Sci.* 2018;101(10):9493–504.
31. Buitenhuis B, Røntved CM, Edwards SM, Ingvarsen KL, Sørensen P. In depth analysis of genes and pathways of the mammary gland involved in the pathogenesis of bovine *Escherichia coli* mastitis. *BMC Genomics.* 2011;12(1):130.
32. Gunther J, Esch K, Poschadel N, Petzl W, Zerbe H, Mitterhuemer S, Blum H, Seyfert HM. Comparative kinetics of *Escherichia coli*- and *Staphylococcus aureus*-specific activation of key immune pathways in mammary epithelial cells demonstrates that *S. aureus* elicits a delayed response dominated by interleukin-6 (IL-6) but not by IL-1A or tumor necrosis factor alpha. *Infect Immun.* 2011;79(2):695–707.
33. Jordan SC, Choi J, Kim I, Wu G, Toyoda M, Shin B, Vo A. Interleukin-6, a cytokine critical to mediation of inflammation, autoimmunity and allograft rejection: therapeutic implications of IL-6 receptor blockade. *Transplantation.* 2017;101(1):32–44.
34. Yang D, Chen Q, Hoover DM, Staley P, Tucker KD, Lubkowski J, Oppenheim JJ. Many chemokines including CCL20/MIP-3a display antimicrobial activity. *J Leuc Biol.* 2003;74(3):448–55.
35. Mitterhuemer S, Petzl W, Krebs S, Mehne D, Klanner A, Wolf E, Zerbe H, Blum H. *Escherichia coli* infection induces distinct local and systemic transcriptome responses in the mammary gland. *BMC Genomics.* 2010;11:138.
36. Zabel BA, Rott A, Butcher EC. Leukocyte chemoattractant receptors in human disease pathogenesis. *Annu Rev Pathol.* 2015;10:51–81.
37. Roe K, Gibot S, Verma S. Triggering receptor expressed on myeloid cells-1 (TREM-1): a new player in antiviral immunity? *Front Microbiol.* 2014;5:627.
38. Meyerholz M, Heimes A, Brodhagen J, Rohmeier L, Eickhoff T, Jander S, Huelsebusch A, Weikard R, Günther J, Seyfert H-M, et al. ChronMast - a model to study functional genetic variation of mastitis susceptibility. Auckland: World congress on genetics applied to livestock production 2018; 2018.
39. Weikard R, Goldammer T, Brunner RM, Kuehn C. Tissue-specific mRNA expression patterns reveal a coordinated metabolic response associated with genetic selection for milk production in cows. *Physiol Genomics.* 2012;44(14):728–39.
40. Weikard R, Goldammer T, Eberlein A, Kuehn C. Novel transcripts discovered by mining genomic DNA from defined regions of bovine chromosome 6. *BMC Genomics.* 2009;10:186.
41. NCBI. BLAST: Basic local alignment search tool. <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
42. NCBI. Primer-BLAST: A tool for finding specific primers. https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=PrimerHome.
43. Hammon HM, Frieten D, Gerbert C, Koch C, Dusel G, Weikard R, Kuhn C. Different milk diets have substantial effects on the jejunal mucosal immune system of pre-weaning calves, as demonstrated by whole transcriptome sequencing. *Sci Rep.* 2018;8(1):1693.
44. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011;17(1):10–2.
45. Andrews S. FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (2010). Accessed 13 Apr 2017.
46. Robinson A. QualityTrim. <https://bitbucket.org/arobinson/qualitytrim> (2015). Accessed 24 Apr 2017.
47. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 2015;12(4):357–60.
48. *Bos taurus* UMD3.1. ftp://ftp.ensembl.org/pub/release-87/fasta/bos_taurus/dna/ (Accessed 21 Mar 2016).
49. Ensembl-Annotation release 87. ftp://ftp.ensembl.org/pub/release-87/gtf/bos_taurus/ (2016).
50. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2014;30(7):923–30.
51. Team RC. R: The R project for statistical computing. <https://www.r-project.org> (2016).

52. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. *Nat Biotechnol*. 2011;29(1):24–6.
53. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol*. 2013;31(1):46–53.
54. QIAGEN. IPA: Ingenuity pathway analysis software. <https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis>. Accessed 17 July 2018.
55. Larsson J. R package eulerr: area-proportional Euler and Venn diagrams with ellipses. <https://cran.r-project.org/web/packages/eulerr/eulerr.pdf> (2018). Accessed 15 Nov 2018.
56. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015;33(3):290–5.
57. GffCompare: Program for processing GTF/GFF files. <https://ccb.jhu.edu/software/stringtie/gffcompare.shtml> Accessed 10 May 2017.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



5 Diskussion

Zur Verbesserung der Empfindlichkeit der RNA-Seq-Analyse gegenüber marginal vorkommender Transkripte bei der Transkriptomanalyse von Milchdrüsenparenchym sollte eine Methode entwickelt werden, welche den Anteil hoch exprimierter Milchproteingen-Transkripte reduziert. Die während der Laktation in bovinem Eutergewebe bis zu 80% vorkommenden stark transkribierten Milchproteingen-Transkripte (CANOVAS et al., 2014; IBEAGHA-AWEMU et al., 2016; BHAT et al., 2019) erschweren das Auffinden und Identifizieren von Transkripten mit niedrigerem Expressionsniveau und machen für eine präzise Detektion eine hohe Sequenziertiefe notwendig. Eine Steigerung der Sequenziertiefe ist wiederum mit hohem Kostenaufwand verbunden (SIMS et al., 2014). Das Depletieren des Anteils stark exprimierter Transkripte soll bei gegebener Sequenziertiefe eine detaillierte Katalogisierung von Transkriptelementen ermöglichen, welche Bedeutung für physiologische und pathologische Mechanismen im Eutergewebe aufweisen. Gleichzeitig sollen so zusätzliche Kosten vermieden werden. Das Verstehen der regulatorischen Abläufe im Euter soll es ermöglichen, Vorgänge zu identifizieren, die im Falle einer Euterinfektion auf Empfänglichkeit und Pathogenese Einfluss nehmen.

5.1 Die Methode

5.1.1 Entwicklung und Optimierung der Methode

Im Vorfeld dieser Studie wurden zur Entwicklung und Optimierung der Methode Vorversuche an depletiertem und nicht-depletiertem Milchdrüsengewebe zweier laktierender Milchkühe mit vergleichender RT-qPCR durchgeführt. Diese Versuche sollten unter anderem den Effekt der Prozedur an sich auf das Probenmaterial durch den Einsatz von lediglich Puffer ohne Oligonukleotide aufzeigen. Außerdem wurden Modifizierungen der Bedingungen während des Hybridisierungsvorganges vorgenommen. Die ursprünglich gewählte Temperatur während der Hybridisierung, also während der Anlagerung eines komplementären Einzelstrangs einer DNA an die Ziel-RNA durch Wasserstoffbrückenbindungen zwischen den jeweils komplementären Basen, wurde von 70°C auf 65°C gesenkt, wobei die Dauer des Vorgangs von 5min konstant blieb. Die Sets der für Casein- und Molkenproteingentranskripte komplementären Oligonukleotide wurden optimiert, und die eingesetzte Konzentration wurde den jeweiligen Mengen der im RNA-Pool befindlichen hoch exprimierten Milchproteingen-Transkripten, wie in der Literatur beschrieben, angepasst (Variante A), so dass die Konzentration auf das Vorhandensein der jeweiligen Casein- und Molkenproteingentranskripte abgestimmt wurde. Dabei wurden mehrere Änderungen der Mengenverhältnisse der Oligonukleotide getestet und verschiedene Varianten von abgeleiteten komplementären Oligonukleotiden überprüft. Gleichzeitig kam zur Evaluierung der Methode ein weiteres Set zum Einsatz, mit gleicher Konzentration an DNA-Oligonukleotiden für alle Casein- und Molkenproteingentranskripte (Variante B) (Konzentrationen beider Varianten in Tabelle 7 der Veröffentlichung Brodhagen et

al., 2019). Schon in den Vorversuchen konnten Reduzierungen der Milchproteingen-Transkripte von 30 bis 90% erreicht werden. Exemplarisch wurden Bibliotheken für die RNA-Seq-Analyse aus denselben Proben, wie schon mittels RT-qPCR analysiert, erstellt und der Effekt der RNase H abhängigen Depletion überwacht. Es war möglich, mittels RT-qPCR und RNA-Seq Milchproteingen-Transkripte zu quantifizieren und den Effekt der Depletion nachzuweisen.

5.1.2 Die Auswahl der Versuchstiere

Die Evaluierung dieser Depletionsmethode erfolgte an bovinem Euterparenchym dreier Holstein-Friesian-Erstkalbinnen. Mit der Wahl der Erstkalbinnen als Versuchstiere sollte ein möglichst einheitliches Probenmaterial in Bezug auf das Milchdrüsentranskriptom gewonnen werden. Es wurde die Voraussetzung des gleichen Alters aller Tiere erfüllt, welche sich zudem im gleichen Stadium der Laktation befanden. Ebenfalls wurde vermieden, dass durch vorherige Laktationen Veränderungen oder Verletzungen an Zitzen oder Euter durch die Anwendung von Melktechnik in früheren Laktationen bestanden. Auch von Bedeutung war, dass die Versuchstiere vor der experimentellen Induktion einer Mastitis an keiner ungeplanten Euterinfektion litten, was bei multiparen Milchkühen ein größeres Risiko darstellt als bei primiparen (BARKEMA et al., 1998). Es ist davon auszugehen, dass eine primipare Milchkuh weniger Pathogenkontakt hatte als eine multipare. Kontakt mit einem bereits bekannten Pathogen könnte zur Änderung der Immunreaktion durch das erworbene Immunsystem führen. Die drei für die Evaluierung der Methode gewählten Tiere zeigten während des Infektionsversuchs, bei dem insgesamt 12 Tiere infiziert wurden, unterschiedlich starke Reaktion auf das Bakterium *E. coli*. Zwar zeigten alle drei Tiere einen Einbruch in der Milchproduktion, doch sank bei Kuh 3 die Produktion mit nur noch 38% am stärksten. Auch ergab nur bei Kuh 3 der pathologische Bericht eine granulozytäre Infiltration in das infizierte Euterviertel. Die Auswahl der Tiere ermöglicht die Evaluierung der entwickelten Methode an infiziertem und nicht-infiziertem Gewebe mit unterschiedlicher lokaler Reaktion auf das Bakterium *E. coli* und unterschiedlicher Beeinträchtigung der Milchleistung. Die Depletion von Milchproteingen-Transkripte für die RNA-Seq wäre beim infizierten Euterviertel der Kuh 3, wo der Anteil an hoch exprimierten Transkripten nur noch 38% ausmacht, theoretisch nicht notwendig gewesen, doch soll auch an diesem Beispiel die Auswirkung der Methode überprüft werden und außerdem eine Vergleichbarkeit der Proben erhalten bleiben.

5.1.3 Die frühe Laktation als Zeitpunkt der Probennahme

Die Wahrscheinlichkeit, im Zeitraum der frühen Laktation an einer klinischen Mastitis zu erkranken, ist für die Milchkuh besonders hoch. Burvenich *et al.* (BURVENICH et al., 2003) begründen die besonders hohe Empfänglichkeit in diesem Zeitraum durch eine Immunschwäche, vermutlich verursacht durch den hohen Energieaufwand zur Erreichung der Milchproduktion und damit verbunden eine geringe Widerstandsfähigkeit gegenüber Stress. Außerdem beschreiben sie eine Akkumulation physiologischer Prozesse als Grund für das vermehrte Auftreten von Euterentzündungen im postpartalen Zeitraum (BURVENICH et al.,

2007). Vangroenweghe *et al.* (VANGROENWEGHE *et al.*, 2005) vermuten, dass die lokalen Hormonschwankungen, welche zu strukturellen Änderungen und zu verändertem Sekretionsstatus im Gewebe führen, die Immunantwort und die klinische Reaktion auf eine Mastitis stören könnten. Die Probennahme erfolgte in dieser Studie in diesem beschriebenen kritischen Zeitfenster der frühen Laktation, an Tag 36 \pm 3,4 nach der Abkalbung. Als sehr wichtig wurde erachtet, dass alle Erstkalbinnen genau im gleichen Stadium der Laktation experimentell infiziert wurden, um die Vergleichbarkeit zwischen den Tieren zu gewährleisten.

5.1.4 Anwendung der Methode an mit *E. coli* infiziertem und nicht-infiziertem bovinem Eutergewebe

Die Anwendung der entwickelten Depletionsmethode zur Entfernung häufig exprimierter Transkripte aus Gesamt-RNA erfolgte an Eutergewebe mit und ohne experimentell induzierter Entzündung. Zum einen ergab sich so die Möglichkeit, die Methode an Gewebe mit unterschiedlich stark ausgeprägter Milchproduktion und somit unterschiedlich hoch exprimierten Anteil an Milchproteingen-Transkripten zu evaluieren und zum anderen ist eine differentielle Transkriptomanalyse von infiziertem vs. nicht-infiziertem Parenchym exemplarisch durchführbar. Die durch die Depletion der stark exprimierten Transkripte verbesserte Empfindlichkeit gegenüber marginal exprimierten Genen soll dabei möglichst zusätzlichen Aufschluss über regulatorisch relevante Prozesse der Entzündung geben.

Zur experimentellen Induktion einer klinischen Euterentzündung wurde das Bakterium *E. coli* 1303 verwendet. Diese ubiquitär vorkommende enterobakterielle Spezies gilt als bedeutende Ursache klinischer Mastitiden (HOGAN & LARRY SMITH, 2003) und das gewählte, aus einer klinischen Mastitis gewonnene Isolat wurde bereits vielfach zur experimentellen Induktion von Euterentzündungen verwendet (MITTERHUEMER *et al.*, 2010; GUNTHER *et al.*, 2011; GUNTHER *et al.*, 2017).

5.1.5 Enzymatischer Abbau hoch exprimierter Transkripte mittels RNase H

Zu den beiden genutzten Hauptstrategien für die Depletion hoch-abundanter Transkripte gehören das Abfangen der Zielsequenzen mittels komplementärer Oligonukleotide und daran gekoppelter paramagnetischer Beads und der Verdau mittels RNase H von doppelsträngigen RNA:DNA-Sequenzen, bestehend aus Zielsequenz und dazu komplementären Oligonukleotiden. Studien, welche den Vergleich zwischen den beiden Strategien ziehen, zeigen, dass beide Varianten signifikante Entfernung der rRNA erreichen und den Anteil hoch exprimierter Transkripte auf unter 20% senken (HERBERT *et al.*, 2018). Trotzdem wird die Variante mit Hilfe des Enzyms RNase H als bessere bei der Anwendung an RNA minderwertigerer Qualität beschrieben und zeigt zusätzlich die einheitlicheren Ergebnisse (ADICONIS *et al.*, 2013; HERBERT *et al.*, 2018). Außerdem ist sie besonders kostengünstig (HUANG *et al.*, 2020). Die gezielte RNA-Entfernung stark transkribierter Transkripte zur Vorbereitung der RNA-Seq wurde erfolgreich in kommerziell genutzten Kits zur Präparation von cDNA-Bibliotheken eingeführt, um globine RNA aus Blutzellen sowie ribosomale und mitochondriale RNA-Fractionen in der

Gesamt-RNA aus verschiedenen Gewebetypen abzubauen (MASTROKOLIAS et al., 2012; BAYATTI et al., 2014; SULTAN et al., 2014; KRJUTSKOV et al., 2016).

Zur Entfernung der häufig vorkommenden Milchproteingen-Transkripte aus dem Pool der RNA von Milchdrüsengewebe wurden für Sequenzen nahe des 3' Endes der Caseingene und Molkenproteingene komplementäre Oligonukleotide abgeleitet. Diese hybridisieren im Pool der Gesamt-RNA an die Zielsequenz und werden anschließend von der doppelstrangspezifischen RNase H abgebaut. Die Zieltranskripte stehen nach Poly(A)-Selektion zum Umschreiben in cDNA und Einbau in Bibliotheken nicht mehr zur Verfügung.

5.1.6 Datenanalyse und Referenzannotation

Nach der RNA-Isolation aus den Euterparenchymproben mit und ohne induzierter Infektion erfolgte die Durchführung der Depletion mittels der Oligonukleotid-Sets der Variante A und B. Zum Vergleich wurden auch Proben ohne Depletion analysiert. Im Anschluss wurde die transkribierte RNA in cDNA umgeschrieben und Bibliotheken generiert. Nach Durchführung der Sequenzierung des Transkriptoms (siehe 2.4) erfolgte die Annotation. Durch das Annotieren werden den abgelesenen Sequenzen Informationen zugeteilt über die Position im Genom, zu Grenzen von Introns und Exons und Namen von Genen und Proteinprodukten. Die Annotation kann entweder auf der Basis einer Referenz, also einer bereits bekannten Beispielsequenz, beruhen oder *de novo* ausgerichtet werden (STEIN, 2001).

Das Zuordnen und Ausrichten der durch die RNA-Seq gewonnenen Sequenzen erfolgte in dieser Studie auf Basis der Ensembl 87 Referenzannotation. Zusätzlich wurden für die Auswertung drei projektspezifische Annotationen erstellt. Hierzu wurden jeweils alle sequenzierten Fragmente der Kontrollgruppe ohne Depletion, der Fragmente aus dem Versuchsansatz „depletiert mit Variante A“ und der Fragmente aus dem Versuchsansatz „depletiert mit Variante B“ ohne Berücksichtigung des Infektionsstatus zu einer separaten Genomannotation auf Transkriptebene assembliert (vereint). Ausgangspunkt der Fragmentassemblierung für die so erstellten drei gruppenspezifischen Annotationen war dabei jeweils die Ensembl 87 Referenzannotation.

Die für dieses Projekt spezifischen, sensitiveren Transkriptom-Annotationen wurden genutzt, um zuvor noch unbekannte Loci (Genorte) zu identifizieren und zu zählen. Dazu erfolgte zunächst ein Abgleich zwischen der gruppenspezifischen Annotation und dem Referenzgenom Ensembl 87 und anschließend der Vergleich zwischen den einzelnen Gruppen. Zusätzlich wurden mit den speziellen Assemblierungen vergleichend die Gesamtzahl von analysierten Loci ermittelt und die Anzahl unter Berücksichtigung des Schwellenwertes FPKM > 1 (siehe Veröffentlichung Brodhagen et al., 2019, Tabelle 5) gezählt.

5.2 Ergebnisse der Evaluierung der Methode

5.2.1 Depletion hochabundanter Milchproteingen-Transkripte aus dem Eutertranskriptom

Das Ziel, hoch exprimierte Transkripte von Genen, kodierend für Casein- und Molkenproteine aus dem Pool der RNA von laktierendem Euterparenchym zu entfernen und dadurch deren Anteil am Gesamt-Transkriptom zu reduzieren, wurde erreicht.

Es ist gelungen mit Hilfe der entwickelten Methode den Anteil an Milchproteingen-Transkripten des RNA-Pools von nicht-infiziertem Eutergewebe mit der experimentellen Variante A von ursprünglich 61% auf 29% (25 – 36%) und mit Variante B auf 24% (21 – 30%) zu reduzieren. Das Transkriptom des mit *E. coli* infizierten Euterparenchyms wies schon vor der Depletion einen geringeren Anteil Milchproteingen-Transkripte auf, weshalb hier sogar eine Dezimierung auf nur noch durchschnittlich 11% (0 - 24%) mit dem Oligonukleotid-Set der Variante A und 9% (0 – 19%) mit Variante B erreicht werden konnte (siehe Veröffentlichung Brodhagen et al., 2019, Abbildung 3).

Trotz der erfolgreichen Reduktion der stark transkribierten Sequenzen ist bei der Betrachtung einzelner Casein- und Molkenproteine aufgefallen, dass die Effizienz der Depletion Variabilität aufzeigt. Nicht alle Transkripte wurden gleichermaßen aus dem RNA-Pool entfernt.

Die erfolgreichste Reduktion konnte bei den Genen CSN1S2 und CSN3 beobachtet werden, wo eine nahezu vollständige Entfernung der Transkripte aus dem RNA-Pool zu verzeichnen war. CSN2 konnte um 75% reduziert werden. Bei CSN1S1, LALBA und PAEP ist die Depletion der Transkripte hingegen nicht oder kaum gelungen.

Um diese unterschiedliche Effizienz der Depletion von den einzelnen Transkripten der Gene kodierend für Milchproteine zu erklären, gibt es unterschiedliche Ansätze und Herangehensweisen:

Eine denkbare Ursache für eine ineffiziente Entfernung der Transkripte könnte das Vorhandensein von internen Poly(A)-Abschnitten mit mehr als 7 aufeinanderfolgenden Adenin Nukleotiden sein, welche den Schritt der Poly(A)-Selektion beeinflussen. Eine visuelle Kontrolle der Sequenzen erlaubt den Ausschluss dieser These.

Unterschiedliche genetische Varianten im Bereich der Sequenz, welche zur Ableitung der Oligonukleotide verwendet wurde, waren anhand dieser Studie mittels Transkriptomanalyse lediglich bei CSN3 zu finden. Eine Varianz in der Sequenz könnte das Hybridisieren der Oligonukleotide und so die Doppelstrangbildung stören und folglich den Verdau durch das Enzym RNase H verhindern und zu verringertem Erfolg bei der Depletion der Zielgene führen. Jedoch konnte CSN3 nahezu vollständig mittels der entwickelten selektiven Depletionsmethode aus dem Pool der RNA entfernt werden, weshalb auch das Vorkommen von Single Nucleotide Polymorphismus (SNPs), also Variationen einzelner Basenpaare, im Bereich der Oligonukleotidsequenz als Ursache differenter Effizienz der Depletion offensichtlich unwahrscheinlich ist.

Auch Unterschiede im Abstand zwischen der abgeleiteten Oligonukleotid-Sequenz und dem 3' Ende des Zielgens können zunächst als mögliche Ursache für eine variable RNA-

Depletionseffizienz ausgeschlossen werden. Ist der Abstand zwischen der abgeleiteten Oligonukleotid-Sequenz und dem Poly(A)-3'Ende des Milchproteingen-Transkripts zu groß, bleibt nach dem Hybridisieren und Spalten des doppelsträngigen Bereichs durch das Enzym RNase H eine dementsprechend lange Restsequenz mit Poly(A)-Schwanz bestehen. Diese ist auch nach der Poly(A)-Selektion für den Einbau in die Sequenzierbibliothek verfügbar und kann womöglich bei späteren Analysen als Milchproteingen-Transkript erkannt werden. Der mittlere Abstand zwischen der Position des komplementären Oligonukleotides und dem 3'Ende aller Zielgene betrug 280 bp (35 – 443 bp). Obwohl die spezifischen Oligonukleotide des CSN2 einen überdurchschnittlich großen Abstand von 284 und 439 bp zum 3'Ende hatten, wurde dieses Casein erfolgreich aus dem Pool der RNA entfernt. Ein angestrebtes, möglichst dem 3'Ende nahes Ableiten der Oligonukleotide war oftmals aufgrund repetitiver Elemente oder durch Variationen in den Sequenzen nicht umsetzbar. Jedoch hatte diese Tatsache in dieser Studie bei CSN2 nachweislich keinen Einfluss auf die RNA-Depletions-Effizienz.

Bei der Ursachenfindung für die unzureichende Entfernung von Transkripten für die Gene LALBA und PAEP konnten nach visueller Kontrolle mittels Integrative Genomics Viewer (IGV) keine genetischen Varianten in der Sequenz der Transkripte ausfindig gemacht werden, welche das Hybridisieren der Oligonukleotide verhindern könnten. Es konnte allerdings bei der Betrachtung depletierter Proben festgestellt werden, dass eine erhöhte Read-Abdeckung am 3'Ende der beiden Transkripte vorliegt. Das heißt, dass vermehrt Lesevorgänge am Poly(A)-Ende der Transkripte nachweisbar waren. Jedoch war diese Abdeckung am 5'Ende der Transkripte nicht zu erkennen. Bei der Inspektion nicht-depletierter Proben konnte dieses Phänomen der Verzerrung nicht beobachtet werden. Daraus lässt sich schlussfolgern, dass die 3'Enden der LALBA und PAEP Transkripte nach dem Hybridisieren und dem Verdau mittels RNase H womöglich noch Poly(A)-Sequenzen beinhalten und im Schritt der Poly(A)-Selektion wahrscheinlich deswegen nicht effektiv aus dem Pool der Gesamt-RNA entfernt werden konnten. Eine effizientere Depletion dieser beiden Transkripte ist womöglich durch eine dem 3'Ende nähere Ableitung der Oligonukleotide möglich und sollte daher in zukünftigen Oligonukleotid-Sets berücksichtigt werden.

Ebenfalls können Defizite bei der Entfernung der hoch exprimierten Transkripte mit der Methode auftreten, wenn die abgeleiteten Sequenzen der Oligonukleotide mögliche alternative Spleiß-Varianten in den Sequenzen der Zielgene nicht abdecken. Spleißen bezeichnet den Vorgang, bei dem Introns (nicht-kodierende Abschnitte) aus dem zunächst im Zellkern während der Transkription entstandenen RNA-Molekül entfernt und Exons (kodierende Abschnitte) miteinander verbunden werden. Es entsteht aus primärer RNA oder prä-mRNA reife mRNA (Abbildung 2). Dabei gibt es verschiedene Alternativen des Verbindens von Exons, sodass aus einer RNA-Sequenz unterschiedliche mRNAs hervorgehen können und folglich bei der anschließenden Translation auch verschiedene Proteine entstehen (Abbildung 2). Mit dem IGV wurden die Sequenzen, die zur Ableitung der Oligonukleotide genutzt wurden, visuell inspiziert. Die geringe Reduzierung der CSN1S1 Transkripte könnte durch das Vorhandensein der hohen

Anzahl an Spleißstellen in der Zielgen-Sequenz erklärt werden, welche das Hybridisieren der Oligonukleotide an diese verhindert. Laut Rinder genom-Annotation ARS-UCD1.2 beim NCBI sind aktuell 25 Varianten des CSN1S1-Transkripts bekannt, welche zum Zeitpunkt der Primerdesign-Entwicklung noch nicht alle veröffentlicht waren. Beide abgeleiteten Oligonukleotid-Sequenzen von CSN1S1 befinden sich an unterschiedlichen Spleißstellen. Das Vorhandensein eines Introns an diesen Stellen ist möglich, was dazu führen könnte, dass das Hybridisieren der Oligonukleotide nicht an alle Zieltranskripte stattfindet und einige hoch exprimierte Transkriptvarianten nicht abgedeckt werden. Das Ableiten weiterer, auf CSN1S1-Transkriptvarianten abzielende Oligonukleotide, würde den Erfolg der Depletion womöglich verbessern.

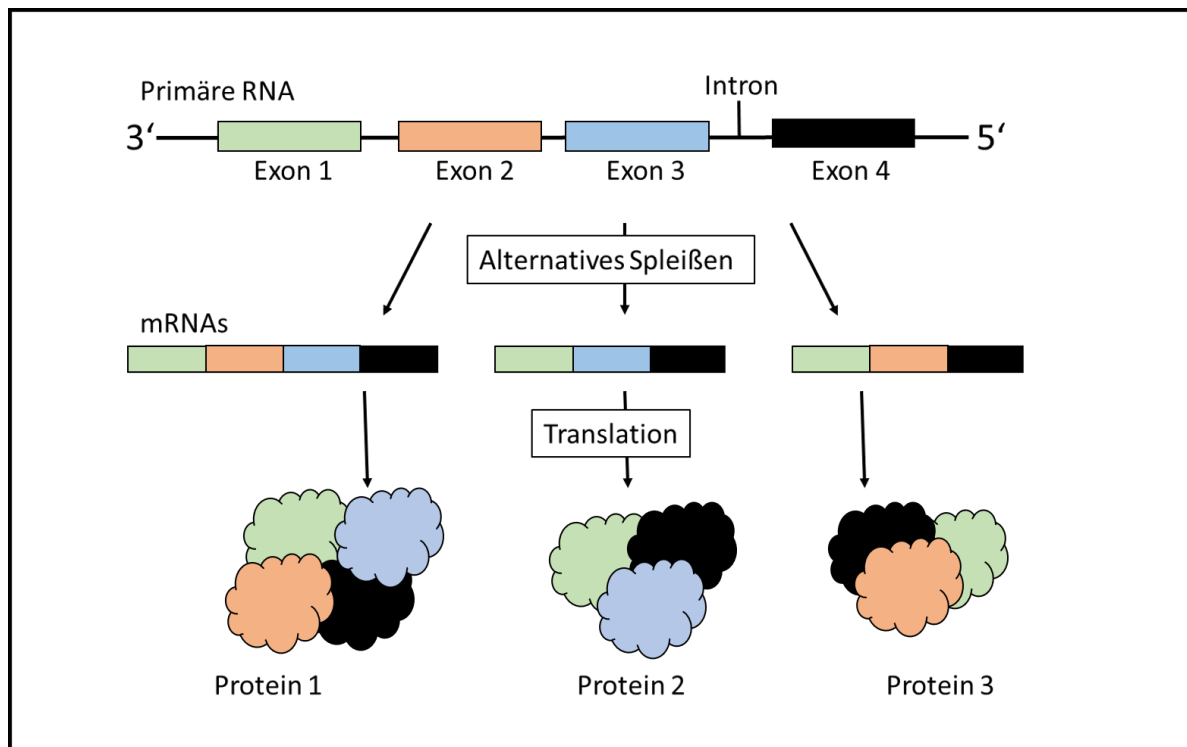


Abbildung 2: Entstehung verschiedener Varianten von mRNA durch alternatives Spleißen von primärer RNA und daraus resultierende Proteinprodukte durch die Translation

Nicht ausgeschlossen werden kann auch, dass Sekundär- und Tertiärstrukturbildung bei den Zielensequenzen dazu führen, dass die Oligonukleotide nicht an den Zielregionen hybridisieren können und deshalb ein Verdau durch die doppelstrangspezifische RNase H nicht stattfindet.

5.2.2 Differenzielle Transkriptomanalyse nach *E. coli*-Infektion von depletierten und nicht-depletierten Milchdrüsengewebe

Was die Ursache für die genetisch vorteilhafte oder unvorteilhafte Ausgangssituation der Milchrinder gegenüber Mastitis ist, ist noch nicht im Detail geklärt. Die verbesserte Empfindlichkeit der RNA-Seq-Analyse gegenüber Transkripten mit niedrigem Expressionslevel

durch die Entfernung der Milchproteingen-Transkripte soll eine Identifizierung der entscheidenden Hintergründe und Ursachen für die unterschiedliche Empfänglichkeit von Euterinfektionen und deren Schwere der Symptomatik erleichtern. In dieser Studie wurde jeweils ein Euterviertel der Holstein-Friesian-Erstkalbinnen 36 +/- 3,4 Tage nach der Abkalbung mit dem Bakterium *E. coli* 1303 experimentell infiziert. 24 Stunden nach der Infektion folgte die Probennahme aus dem infizierten und jeweils einem unbehandelten Euterviertel als nicht-infizierte Kontrolle (Abbildung 3). Eine differentielle Expressionsanalyse von Proben mit provozierte Infektion vs. Kontrolle ohne Anwendung der Depletion wie auch an Proben depletiert mit Variante A und B, sollte signifikant differentiell exprimierte Transkripte aufzeigen und die angestrebte Verbesserung der Detektion von Expressionsunterschieden durch die Depletionsmethode darstellen.

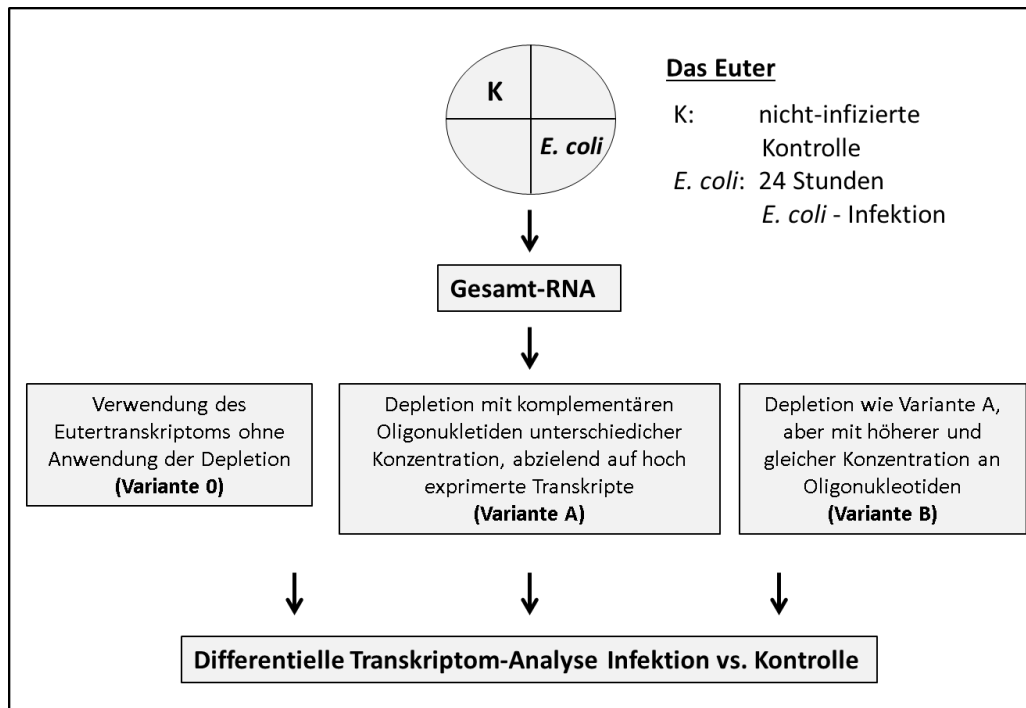


Abbildung 3: Differentielle Transkriptomanalyse von Euterparenchym ohne die Anwendung der Depletion und mit Anwendung der Depletion nach Variante A oder B

In dieser Studie wies das Transkriptom des Parenchyms der nicht-infizierten Euterviertel durchschnittlich 61% (52 – 71%), das mit *E. coli* experimentell infizierte Gewebe hingegen nur durchschnittlich 30% (0,6 – 57%) Milchproteingen-Transkripte auf (Veröffentlichung Brodhagen et al., 2019 Abbildung 3).

Der Rückgang der Milchproduktion nach Euterinfektion mit dem Umgebungskeim *E. coli* wurde bereits in der Literatur beschrieben (BURVENICH et al., 2003; HAGIWARA et al., 2014) und kann durch die womöglich entstandenen massiven Gewebeschäden im Parenchym erklärt werden. Burvenich et al. (BURVENICH et al., 2003) erklären die entstehenden Schäden am Epithel der

Milchdrüse durch das Eindringen der Bakterien und der Fresszellen, welche vom Abwehrsystem des Wirts ausgesendet werden. Insbesondere Neutrophile seien notwendig, um die eingedrungenen Bakterien zu eliminieren, doch gerade die von diesen Granulozyten produzierten chemischen Verbindungen (ROS) schädigen die Sekretionszellen der Milchdrüse und verringern so die Milchleistung, genauso wie die Substanzen, die vom Erreger selbst abgesondert werden.

Auch denkbar wäre die Begründung, dass durch die im Vordergrund stehende Immunabwehr im Zeitraum der Infektion und damit verbundene in den Hintergrund tretende Produktion der Milch diese Gene weniger stark transkribiert werden.

Ein Rückgang der exprimierten Milchproteingene in infiziertem Eutergewebe im Gegensatz zu nicht-infiziertem war ebenfalls in nach dieser Methode depletierten Proben nachzuweisen.

Nachfolgend soll auf die Ergebnisse der differentiellen Transkriptomanalyse eingegangen werden. Die Auswertung der Analysen von Euterparenchymproben ohne Depletion zeigte im Vergleich Infektion vs. Kontrolle 1079 signifikant differentiell exprimierte Gene. Das gezielte Entfernen der Milchproteingen-Transkripte führte zu einer Steigerung der signifikant unterschiedlich exprimierten Gene auf 1290 mit der Variante A und 1226 mit Variante B, was für eine verbesserte Sensitivität der RNA-Seq-Analyse spricht.

Trotz der Anwendung der gezielten RNA-Depletionsmethode sollten die allgemeinen physiologischen Reaktionsmuster in der Milchdrüse während der Infektion mit *E. coli* nicht beeinträchtigt werden. Betrachtet wurden deshalb signifikant unterschiedlich exprimierte biologische Signalwege, basierend auf den mittels RNA-Seq analysierten signifikant differentiell exprimierten Loci. Festgestellt wurde, dass beinahe alle analysierten Signalwege (146 Stück) in depletierten wie auch nicht-depletierten Proben nachgewiesen wurden. Lediglich eine geringe Anzahl an Signaltransduktionspfaden konnte in den depletierten Proben der Variante A (15) und in Variante B (19) im Gegensatz zu den unbehandelten Proben nicht identifiziert werden. Jedoch war es auch möglich mit beiden Varianten der Depletion zusätzliche Signaltransduktionswege (15 bei Variante A und 14 bei Variante B) aufzudecken. Daraus lässt sich ableiten, dass das Verfahren der RNA-Depletion nicht zu einer unbeabsichtigten Verzerrung der globalen biologischen Signalwege führt.

Exemplarisch wurden Gene betrachtet, welche nachweislich eine Bedeutung für die Infektion mit *E. coli* haben und welche für Zytokine, Chemokine und entzündungsrelevante Rezeptoren kodieren (siehe Veröffentlichung Brodhagen et al., 2019, Tabelle 4). Es ließ sich feststellen, dass die Methode der gezielten Depletion die Sensitivität der RNA-Seq-Analyse gegenüber einigen dieser durch die *E. coli* - Infektion aktivierten Gene verbessert. Hauptregulatoren der Akute-Phase-Reaktion, das IL6-Gen (GUNTHER et al., 2011; JORDAN et al., 2017) und CCL16-Gen, konnten mittels der differentiellen Transkriptomanalyse (Infektion vs. Kontrolle) lediglich in depletierten Proben als signifikant unterschiedlich exprimiert identifiziert werden. Auch signifikant differentielle Expression des CCL20-Gens (YANG et al., 2003; MITTERHUEMER et al.,

2010; ZABEL et al., 2015) konnte ausschließlich in Proben depletiert nach Variante B aufgezeigt werden. Genauso ist es bei relevanten Transmembranrezeptoren. TREM1, welcher auf myeloischen Zellen exprimiert wird, verschiedenste zelluläre Funktionen reguliert und unter anderem durch Neutrophile und Monozyten vermittelte Entzündungsreaktionen verstärkt (ROE et al., 2014), wurde ausschließlich in Proben nach Depletion als signifikant exprimiert analysiert. Auch beim CCR2- und CCR7-Gen, die beide für G-Protein-gekoppelte Rezeptoren kodieren und für die Rekrutierung von Leukozyten wesentlicher Bedeutung sind (ZABEL et al., 2015), konnte eine auf die Infektion bezogene signifikant unterschiedliche Expression nur nach Depletion beobachtet werden. Die durch das Entfernen des massiven Anteils an Milchproteingen-Transkripten verbesserte Sensitivität der RNA-Seq-Analyse ermöglicht bei gleicher Sequenztiefe offensichtlich eine präzisere Identifizierung von Genen mit Relevanz für regulatorische Vorgänge im Euter.

Zur Bestätigung dieser Ergebnisse wurde eine differentielle RT-qPCR-Analyse für die IL6- und CCL20-Gene zwischen infizierten und nicht-infizierten Gewebeproben durchgeführt. Eine erhöhte Sensitivität gegenüber den beiden Genen wurde in der Depletionsvariante B bestätigt.

Petzl *et al.* (PETZL et al., 2008) beschreiben die Toll-like-Rezeptoren TLR2 und TLR4 24 h nach experimenteller Infektion des bovinen Euters mit *E. coli* als signifikant differentiell exprimiert. Toll-like-Rezeptoren dienen der Erkennung von Pathogenen und spielen somit eine wichtige Rolle im angeborenen Immunsystem. Die Analysen dieser Arbeit zeigen nur für TLR2 signifikant differentielle Expression zwischen infiziertem und nicht-infiziertem Gewebe sowohl in depletierten und nicht-depletierten Proben. TLR3 und TLR6 wurden weder in der Studie von Petzl *et al.* noch in dieser Studie als signifikant exprimiert gefunden. Die unterschiedlichen Ergebnisse zur Expression von TLR4 sind möglicherweise durch den Zeitpunkt der Infektion zu erklären, welcher bei Petzl *et al.* die Mitte und in dieser Studie der Anfang der Laktation war.

Li *et al.* (LI et al., 2019) fassen drei Studien von Mitterhuemer *et al.* (MITTERHUEMER et al., 2010), Sipka *et al.* (SIPKA et al., 2014) und Buitenhuis *et al.* (BUITENHUIS et al., 2011) mit Microarray-Analysen zusammen und nennen 10 am stärksten signifikant exprimierte Gene nach experimenteller Euterinfektion mit *E. coli*. Auch in dieser Studie wurden von diesen genannten Genen IL1B, ICAM1 und TLR2 in nicht-depletierten wie auch in depletierten Proben als signifikant different exprimiert analysiert. MMP9 wurde lediglich in Proben depletiert nach Variante B als signifikant hochreguliert identifiziert. Ein signifikanter Anstieg von IL6 konnten nur in depletierten Proben (Variante A und B) festgestellt werden, nicht aber in Proben ohne Depletion. CXCL8, CXCL2, IL8RB, CXCL6 und GRO1, welche auch unter den 10 am stärksten signifikant unterschiedlich exprimierten Genen der drei Studien waren, wurden in dieser Studie in keiner methodischen Variante als signifikant verändert gefunden. Als Ursache für diese Unterschiede der Ergebnisse wäre denkbar, dass, wie in dieser Studie, im Versuch von Mitterhuemer *et al.* und Sipka *et al.* die Probennahme nach der Tötung der Tiere erfolgte, Buitenhuis *et al.* nutzten eine Biopsiepistole um Proben am lebenden Tier zu nehmen.

Zusätzlich wurden zwar in allen drei Studien die Proben des Euterparenchyms 24 Stunden nach der Infektion genommen, doch befanden sich die Tiere zum Zeitpunkt des Infektionsversuchs in unterschiedlichen Laktationsstadien, was eine Ursache für die Differenzen der Ergebnisse sein könnte. Mitterhuemer *et al.* beginnen den Versuch während der Mitte der Laktation (drei bis sechs Monate *post partum*) und auch Sipka *et al.* beschreiben den Zeitpunkt der mittleren Laktation für ihren Versuch. Buitenhuis *et al.* hingegen induzierten schon vier bis sechs Wochen nach der Abkalbung eine experimentelle *E.coli* Mastitis, was dem Zeitpunkt dieser Studie an Tag 36 +/- 3,4 *post partum* ungefähr entspricht. Bei den Ergebnissen ist die geringe Tierzahl dieser Studie zu beachten. Zur Evaluierung dieser Ergebnisse ist die Auswertung weiterer Transkriptomanalysen einer größeren Anzahl von experimentell infizierten Milchrindern notwendig.

5.2.3 Identifizierung von Transkripten mit niedrigerem Expressionsniveau durch sensitivere RNA-Seq-Analyse nach der Depletion

Mit der Entfernung der hoch exprimierten Transkripte soll eine Verbesserung der Empfindlichkeit der RNA-Seq gegenüber marginal transkribierten Genen erreicht werden.

Zur Beobachtung der Expressionsdaten von Transkripten mit niedrigerem Expressionslevel im Verlauf der Anwendung dieser Methode und somit der Beurteilung der Sensitivität der RNA-Seq-Analyse wurden Expressionsquartile der Transkripte mittels analysierter FPKM-Werte berechnet. Mit den Quartilgrenzen wurden vier Kategorien, abgestuft nach dem Transkriptionsniveau, gebildet und jeweils ein durchschnittlicher FPKM-Werte für jede dieser Kategorien berechnet und zwischen den Proben ohne und mit Depletion nach Variante A und B verglichen. Die Abstufung der Transkriptionsniveaus reichte von sehr niedrig über niedrig und mittel bis hin zu hoch. (siehe Veröffentlichung Brodhagen et al., 2019, Abbildung 6). Der ursprünglich berechnete durchschnittliche FPKM-Wert der einzelnen Expressionskategorien verschob sich durch die Anwendung der Depletionsmethode der Variante A und noch etwas mehr durch Variante B nach oben. So ist davon auszugehen, dass der durchschnittliche FPKM-Wert einzelner Transkripte nach Depletion höher ist als ohne Anwendung der Methode. Eine Detektion dieser Transkripte bei gegebener Sequenziertiefe mittels RNA-Seq wird so also erleichtert.

Um diese Aussage zu unterstützen wurde mittels der Referenzannotation Ensembl 87 und gezählter Reads die Anzahl der Gene, die den Schwellenwert FPKM > 1 in nicht-depletierten Proben, in Proben nach Depletion mit Variante A und depletiert nach Variante B überschritten, analysiert (siehe Veröffentlichung Brodhagen et al., 2019, Tabelle 3). Zusätzlich wurde festgestellt, wie viele Gene ausschließlich nach der Depletion den Schwellenwert erreichten und welche Anzahl an Genen den Schwellenwert nach der Depletion nicht mehr erreichten, es ohne Depletion aber taten. Die Depletion mit Einsatz des Oligonukleotid-Sets der Variante B ergab dabei die größte Zahl an Genen mit einem FPKM > 1, gefolgt von der Anzahl der Gene aus dem Depletionsansatz der Variante A. Im nicht-vorbehandelten Parenchym wurden am

wenigsten Gene mit einem FPKM-Schwellenwert von > 1 ermittelt. Ebenfalls konnte mit dem Depletionsansatz der Variante B die größte Anzahl an Genen identifiziert werden, die erst nach der Depletion den Schwellenwert überstiegen. Mit dieser Variante wurden am wenigsten Gene gezählt, die den FPKM-Wert 1 nach der Anwendung nicht mehr erreichten, es aber in nicht-depletierten Proben taten. Auch diese Ergebnisse bestätigen die Annahme, dass nach der Anwendung der entwickelten Methode zur Entfernung der stark exprimierten Milchproteingen-Transkripte ein Anstieg der durchschnittlichen FPKM-Werte der im RNA-Pool verbliebenen Transkripte erfolgt. Eine Detektion von Transkripten mittels RNA-Seq wird so erleichtert.

Gleiche Tendenzen konnten mittels Korrelationsanalysen in Dichte-Plots abgebildet werden (siehe Veröffentlichung Brodhagen et al., 2019, Abbildung 5). Dargestellt wurden FPKM-Daten der quantitativen Genexpression aus demselben Euterviertel mit und ohne Anwendung der Depletion in *E. coli* infiziertem und nicht-infiziertem Gewebe. Es kann nach der Depletion der hoch exprimierten Transkripte eine Verschiebung der FPKM-Werte der übrigen Transkripte zu höheren Werten beobachtet werden. Es kommt sonst zu keinen Änderungen der Genquantifizierung, was beweist, dass eine systematische Verzerrung der Genexpression auszuschließen ist (Zusatzdatei 2 und 3 der Publikation).

5.2.4 Einsatz der verbesserten Empfindlichkeit der RNA-Seq-Analyse zur Detektion noch unbekannter Loci

Bei einer erreichten höheren Sensitivität der RNA-Seq-Analyse gegenüber geringer transkribierter Gene durch das Entfernen stark exprimierter Transkripte wäre auch eine verbesserte Detektion noch unbekannter Loci zu erwarten. Zum Beweis dieser Annahme wurden spezifische Transkriptomannotationen für jede experimentelle Gruppe (0, A und B), ohne Berücksichtigung des Infektionsstatus, getrennt assembliert (siehe 5.1.5). Zunächst wurde die Gesamtzahl aller Loci ermittelt, welche nach Anwendung der entwickelten gezielten Depletionsmethode höher ausfiel als in Proben ohne Vorbehandlung. Auch unter Berücksichtigung eines Schwellenwertes von FPKM > 1 war die Anzahl bedeutend höher. Ebenfalls ergaben die Analysen dieses Ergebnis für noch nicht annotierte Loci. Depletierte Proben wiesen eine höhere Anzahl nicht annotierter Gene auf im Vergleich zu Proben ohne Vorbehandlung. Die Anwendung der entwickelten Methode zur Depletion von Milchproteingen-Transkripten erleichtert somit durch die verbesserte Empfindlichkeit der RNA-Seq nachweislich das Auffinden neuer, noch unbekannter Loci (siehe Veröffentlichung Brodhagen et al., 2019, Tabelle 5).

5.2.5 Vergleich der experimentellen Varianten A und B in Bezug auf Depletionseffektivität, Verzerrung der Ergebnisse und Verbesserung der Sensitivität der RNA-Seq-Analyse

Zur Depletion der Milchproteingen-Transkripte aus dem Pool der RNA des Euterparenchyms kamen zwei Oligonukleotid-Sets zum Einsatz. Die Konzentration der eingesetzten komplementären Oligonukleotide, abzielend auf die Sequenzen nahe des 3'Endes der zu entfernenden Casein- und Molkenprotein-Transkripte, war in der Variante A den aus der

Literatur bekannten Mengen der Zielgentranskripte angepasst. In Variante B kamen alle Oligonukleotide in gleicher Konzentration zum Einsatz (siehe Veröffentlichung Brodhagen et al., 2019, Tabelle 7).

Die stärkere Reduktion des Anteils an Milchproteingen-Transkripten im Gesamt-Transkriptom des Eutergewebes mit induzierter Infektion mittels *E. coli* und ohne Infektion wurde beim Einsatz der Variante B sichtbar (siehe Veröffentlichung Brodhagen et al., 2019, Abbildung 3). Auch beim Analysieren der spezifischen Expressionsdaten einzelner Casein- und Molkenproteine zeigte das Oligonukleotidset der Variante B die effektivere Reduzierung der Transkripte. Ebenfalls war die Verschiebung nach oben der durchschnittlichen FPKM-Werte der gebildeten Quartile nach der Depletion mit Variante B am größten (siehe Veröffentlichung Brodhagen et al., 2019, Abbildung 6) und es wurde die größte Anzahl an Genen mit Schwellenwert $\text{FPKM} > 1$ gefunden (siehe Veröffentlichung Brodhagen et al., 2019, Tabelle 3).

Allerdings ergaben differentielle Transkriptomanalysen von Expressionsdaten der Kontrollgewebe vs. infiziertem Gewebe mehr signifikant unterschiedlich exprimierte Transkripte mit der Variante A. Auch wurden nach der Anwendung der Variante A mehr Loci nachgewiesen (sowohl mit als auch ohne Anlegen des Grenzwertes von $\text{FPKM} > 1$) wenn die sensitivere, gruppenspezifische Annotation die Grundlage für die Transkriptidentifizierung bildete (siehe 5.1.5). Ebenfalls war die Anzahl bisher noch unbekannter Loci, analysiert mittels projektspezifischer Assemblierung, bei Depletion nach Variante A am größten.

Analysen der Signaltransduktionswege anhand der differentiell exprimierten Transkripte ergaben mit beiden Varianten der Depletion keine Verzerrung der pathophysiologisch ablaufenden Vorgänge während der Euterinfektion. Zusätzlich konnte durch Dichteplots der Korrelationsanalysen von FPKM-Werten zwischen depletierten und nicht-depletierten Proben zwar eine Verschiebung der quantitativen Genexpression zu höheren Werten beobachtet werden, doch die Verteilung blieb gleich.

Obwohl beide Oligonukleotid-Sets (Variante A und B) jeweils Vor- und Nachteile aufweisen, fiel die Entscheidung auf den Einsatz der Variante B zur Depletion des Anteils an stark exprimierten Milchproteingen-Transkripten aus Eutereparenchym in weiterführenden Versuchen. Mit dieser Variante kann die bestmögliche Reduktion der Transkripte erreicht werden.

5.3 Anwendung der Methode zur Identifizierung vorteilhafter Genotypen

Die Notwendigkeit, mehr über den Hintergrund der Empfänglichkeit einer Milchkuh gegenüber Mastitis zu erfahren, erklärt sich durch die Häufigkeit (HALASA et al., 2007) und die wirtschaftlichen Verluste durch Kosten für Prophylaxe und Behandlung, Milchertragsrückgang und Tierabgänge, welche mit dieser Infektion in Verbindung stehen. Euterentzündungen gelten in der konventionellen Milchviehhaltung als die am häufigsten auftretenden Erkrankungen (POL & RUEGG, 2007), oft in Verbindung mit einer Beeinflussung des Tierwohls, weshalb auch gesellschaftspolitisch eine große Bedeutung besteht (BURVENICH et al., 2003; OLIVEIRA et al., 2013). Doch auch der wiederkehrende Einsatz von Antibiotika zur Bekämpfung auftretender

Mastitiden in der konventionellen Milchviehhaltung und die damit in Verbindung stehende Förderung von resistenten Bakterienpopulationen erfordern eine Minimierung der Krankheitsinzidenz in Milchkuhherden (TEUBER, 2001; WALLMANN & HEBERER, 2014).

Zur Prävention der Ausbreitung ansteckender Mastitiden in Milchviehbetrieben wurde eine Strategie entwickelt. Diese besteht aus dem Dippen der Zitze nach dem Melken, der Anwendung einer antibiotischen Therapie zum Trockenstellen einer Milchkuh am Ende jeder Laktation, der richtigen Behandlung klinisch erkrankter Tiere, der Keulung chronisch kranker Tiere und der Wartung und hygienischen Reinigung der Melkwerkzeuge (RUEGG, 2017). Neben der Einhaltung dieser Strategie kommt eine gezielte Selektion in Hinblick auf die Resistenz der Milchkühe gegenüber Mastitis in der Rinderzucht zum Einsatz.

Zusätzlich zur schon seit langem eingesetzten phänotypischen Selektion auf Milchleistung, Melkbarkeit und der somatischen Zellzahl als indirekter Parameter für die Mastitisempfindlichkeit (RUPP & BOICHARD, 2003; KUEHN et al., 2008; SENDER et al., 2013) wird zunehmend mit Hilfe genomischer Informationen selektiert. Jedoch fehlt es in diesem Bereich noch an detaillierter Forschung und zusätzlichen Erkenntnissen. Erfolgreich eingesetzt wurden bereits markerassoziierte Selektionen, welche auf Informationen über Quantitative Trait Loci (QTL) basieren (KUEHN et al., 2008). Immer mehr Studien veröffentlichen QTL, welche die Empfindlichkeit und Pathogenese der Mastitis in Rindern beeinflussen (KUEHN et al., 2003; BRAND et al., 2009; MEREDITH et al., 2013; WANG et al., 2015; MEYERHOLZ et al., 2018; KURZ et al., 2019). Eine Datenbank mit Kandidatengenomen und genetischen Markern für Milchproduktion und Mastitis führt Ergebnisse aus unterschiedlichen Studien zu diesem Thema zusammen (OGOREVC et al., 2009).

Um den Fortschritt in diesem Mastitisempfindlichkeit erforschenden Bereich voran zu treiben, QTL besser identifizieren zu können und Merkmale zu identifizieren, welche eine bessere Phänotypisierung von Kühen erlauben, sind weitere Untersuchungen der Hintergründe regulatorischer Abläufe im Euter notwendig. Transkriptomanalysen von Eutergewebe können dazu beitragen, pathophysiologische Vorgänge zu verstehen. Die entwickelte Methode wiederum ermöglicht es, die Untersuchung des Eutertranskriptoms mittels RNA-Seq günstiger und präziser zu gestalten.

Studien zur Untersuchung von gegenüber Euterinfektionen genetisch vorteilhafter und unvorteilhafter Genotypen (MEYERHOLZ et al., 2018) erfordern die Anwendbarkeit einer Methode zur tiefen Phänotypisierung, z.B. anhand des Transkriptoms, an infiziertem wie auch nicht-infiziertem bovinem Eutergewebe verschiedener Laktationsstadien. Ergebnisse dieser Studie zeigen, dass die Anwendung der entwickelten gezielten Depletionsmethode an Gewebe mit einer *E. coli*-induzierten Infektion unterschiedlichen Ausmaßes möglich ist. Unabhängig vom Status der lokalen Entzündungsreaktion und dem Rückgang der Milchproduktion, verbunden mit der Minimierung der Menge an Milchproteingen-Transkripten, zeigt die entwickelte Methode zur Reduzierung häufig exprimierter Casein- und Molkenproteintranskripte

zuverlässige Ergebnisse in Form einer verbesserten Sensitivität der RNA-Seq-Analyse ohne Verzerrung der biologischen Reaktionsmuster. Denkbar ist daher ebenfalls eine Anwendung an Euterparenchym anderer Laktationsstadien und an Gewebe infiziert mit anderen Keimen als in diesem Versuchsaufbau getestet. So ergibt sich ein breites Spektrum des Einsatzes dieser Methode in zukünftigen Studien.

6 Schlussfolgerung und Ausblick

Die Entwicklung und Evaluierung einer auf hoch exprimierte, für Transkripte von Milchprotein kodierenden Genen abzielende Depletionsmethode für die differentielle Transkriptomanalyse von bovinem Milchdrüsenparenchym ist erfolgreich gelungen. Die während der Laktation im bovinen Eutertranskriptom in großer Menge vorkommenden Transkripte der Gene, kodierend für Casein- und Molkenproteine, konnten reduziert werden. Außerdem konnte die angestrebte verbesserte Sensitivität der RNA-Seq-Analyse gegenüber Genen niedrigerem Expressionsniveaus anhand verbesserter Identifizierung differentiell exprimierter und neuer unbekannter Transkripte und zusätzlich durch den Anstieg von FPKM-Werten der im Pool der RNA verbliebenen Gene nachgewiesen werden. Trotz der Anwendung dieser entwickelten Methode kam es zu keinen unerwünschten Verzerrungen in der quantitativen Genexpression oder in biologischen Signaltransduktionsmustern. Die Depletionseffizienz der experimentell getesteten Varianten A und B, mit jeweils unterschiedlicher Konzentration an für die Zielgene komplementären Oligonukleotiden, unterschied sich nur geringfügig. Beide Oligonukleotid-Sets (A und B) wiesen minimale Vor- und Nachteile auf. Für den zukünftigen Einsatz der gezielten Depletion in bovinem Euterparenchym fiel die Entscheidung aufgrund der geringfügig besseren Reduktion der Milchproteingen-Transkripte auf die Variante B.

Im Rahmen des Verbundprojekts „ChronMast“ wurde die experimentelle Variante B der gezielten Depletion am Eutertranskriptom von *E. coli*- oder *S. aureus*-infizierten und nicht-infizierten Eutervierteln 36 vorausgewählter Erstkalbinnen angewendet (MEYERHOLZ et al., 2018). Die ausgewählten Färsen wurden genotypisch selektiert und unterschieden sich in ihrer Empfänglichkeit gegenüber Mastitiden. Die Anwendung der auf stark exprimierte Milchproteingen-Transkripte abzielenden Depletionsmethode mit der verbundenen verbesserten Sensitivität der RNA-Seq sollte die Identifizierung regulatorischer, die Empfänglichkeit gegenüber Mastitis beeinflussender Abläufe und Signalwege im bovinen Eutertranskriptom erleichtern. So soll es zukünftig möglich sein, den genetischen Hintergrund der für Mastitisempfindlichkeit ursächlichen, pathophysiologischen Abläufe detaillierter aufzudecken. Durch die noch ausstehende Analyse und Auswertung der gewonnenen genomischen Daten soll möglichst ein Identifizieren phänotypischer Merkmale, die in Verbindung mit der Mastitisempfindlichkeit stehen, ermöglicht werden.

Abschließendes Ziel ist eine Etablierung stabiler, leicht identifizierbarer Biomarker. Möglich wären dafür genomische Merkmale, wie SNPs oder aber phänotypische Besonderheiten, welche schnell und kostengünstig nachweisbar sind. Diese Marker wären prognostisch nutzbar, könnten einen praxisnahen Einsatz ermöglichen und entweder für die Selektion und Zucht resistenter Milchrinderpopulationen verwendet werden oder bei der Entscheidung über Managementmaßnahmen helfen, wie z.B. über die Notwendigkeit einer prophylaktischen, antibakteriellen Behandlung während des Trockenstellens.

7 Zusammenfassung

Wirtschaftliche Verluste durch Kosten für die Prophylaxe, Behandlung erkrankter Tiere, Tierverluste und Einbußen in der Milchproduktion sind Folgen von Euterentzündungen, welche als am häufigsten auftretende Erkrankungen in Milchviehbetrieben gelten (SEEGERS et al., 2003; HALASA et al., 2007; HUIJPS et al., 2008; PIEPERS et al., 2009; FÁVERO et al., 2015; ROLLIN et al., 2015; VAN SOEST et al., 2016). Die Pathogenese und Inzidenz dieser Faktorenkrankheit wird durch die Umgebung, das Pathogen und durch den Wirt beeinflusst (BURVENICH et al., 2003). Neben der gesellschaftspolitischen Bedeutung aufgrund der Tierwohlbeeinträchtigung (BURVENICH et al., 2003) ist auch der steigende Antibiotikaeinsatz, verbunden mit zunehmenden Resistenzen der Bakterienpopulationen (TEUBER, 2001; WALLMANN & HEBERER, 2014), ein Grund für die Zucht widerstandsfähiger Milchrinderpopulationen mit Resistenz gegenüber Mastitiden.

Zunehmend wird in der Zucht von Milchrindern nicht nur auf phänotypische Eigenschaften wie Melkbarkeit, Leistung, Fertilität und somatische Zellzahl in der Milch selektiert, sondern auch genomische Informationen werden einbezogen (KUEHN et al., 2008). Um mehr über den genetischen Hintergrund und Zusammenhang von molekularen Mechanismen und pathophysiologischen Vorgängen der Empfänglichkeit gegenüber Euterinfektionen zu erfahren, wurden und werden RNA-Seq-Analysen des bovinen Eutertranskriptoms durchgeführt.

Das besonders häufige Auftreten von klinischen Euterentzündungen postpartum, im Zeitraum der frühen Laktation, macht eine Probennahme und Anwendung der RNA-Seq zu diesem Zeitpunkt notwendig (BURVENICH et al., 2003; VANGROENWEGHE et al., 2005; BURVENICH et al., 2007). Jedoch besteht das Eutertranskriptom während der Laktation bis zu 80% (CANOVAS et al., 2014; IBEAGHA-AWEMU et al., 2016; BHAT et al., 2019) aus Transkripten von Genen kodierend für Casein- (CSN1S1, CSN1S2, CSN2 und CSN3) und Molkenproteine (PAEP und LALBA), was eine Identifizierung und Analyse von Genen mit niedrigerem Expressionslevel erschwert. Um eine erhöhte Sequenziertiefe verbunden mit gesteigertem Kosteneinsatz zu umgehen und trotzdem eine präzise Identifizierung weniger stark exprimierter, jedoch für die Regulationsvorgänge im Euter relevanter Transkripte zu ermöglichen (BARTEL, 2004; HE & HANNON, 2004; LAWLESS et al., 2013; YANG et al., 2018), muss der Anteil stark transkribierter Gene im Transkriptom des Euterparenchyms reduziert werden.

Es wurde darum eine Methode zur Depletion hoch abundanter Milchproteingen-Transkripte aus dem Transkriptom von bovinem Euterparenchym entwickelt und evaluiert. In dieser Studie wurde das Eutergewebe dreier Holstein-Friesian-Erstkalbinnen, aus jeweils einem Eutervierviertel mit einer *E. coli*-Infektion und einem Kontrollvierviertel genutzt. Zur gezielten Entfernung der hoch exprimierten Zieltranskripte kamen Oligonukleotide, komplementär für Sequenzen nahe des 3'Endes dieser Transkripte zum Einsatz. Nach der Hybridisierung der DNA-Nukleotide an die Ziel-RNA wurde eine doppelstrangspezifische RNase H zum Verdau verwendet. Folglich kam es zur Deadenylierung der Milchproteingen-Transkripte. Im anschließenden Schritt der Poly(A)-

Selektion wurden diese Transkripte aus dem Pool der Gesamt-RNA entfernt und standen für eine anschließende Umschreibung in cDNA-Bibliotheken und differentielle RNA-Seq-Analyse nicht mehr zur Verfügung.

Mit dieser entwickelten Methode ist es gelungen, den Anteil von Milchproteingen-Transkripten in den Proben von durchschnittlich 61% in Eutervierteln ohne Infektion auf bis zu durchschnittlich 24% zu reduzieren. Infiziertes Parenchym wies durchschnittlich 30% stark exprimierte Transkripte auf, was durch die Anwendung der Depletion im Durchschnitt auf bis zu 9% reduziert werden konnte. Zur Bestätigung der verbesserten Sensitivität der RNA-Seq-Analyse gegenüber niedrig exprimierten Transkripten wurde ein höherer Anteil an signifikant differentiell exprimierten Transkripten beim Vergleich von *E. coli* infiziertem und nicht-infiziertem Euterviertel durch die Anwendung der Depletionsmethode nachgewiesen. Zusätzlich konnten mehr unbekannte Loci detektiert werden und die durchschnittlichen FPKM-Werte der analysierten Gene stiegen an. Trotzdem kam es zu keinen unerwünschten Verzerrungen in der quantitativen Genexpression oder in biologischen Signaltransduktionsmustern.

Die Entwicklung und Evaluierung einer gezielten Depletionsmethode zur Reduzierung des Anteils von stark exprimierten Milchproteingen-Transkripten im Eutertranskriptom ist gelungen und kann in zukünftigen Studien zur Verbesserung der Sensitivität der RNA-Seq-Analyse gegenüber weniger stark transkribierten Genen angewendet werden. Das präzise Detektieren regulatorisch relevanter Gene bei gegebener Sequenziertiefe erleichtert so das Verstehen genetischer Hintergründe der Empfänglichkeit von Milchrindern gegenüber Euterinfektionen. Es ist zu erwarten, dass eine präzisere und kostensparende Analyse des Transkriptoms dazu beiträgt, die Identifizierung von mit Mastitisempfindlichkeit in Verbindung stehenden phänotypischen Merkmalen voranzutreiben. Dies würde eine Identifizierung und Etablierung stabiler Biomarker erleichtern, die die Zucht auf Mastitis resistente Rinderpopulationen unterstützen.

8 Summary

Economic losses due to costs for prophylaxis, treatment of infected animals, animal losses and reduced milk production are consequences of mastitis, which is considered the most common disease on dairy farms (SEEGERS et al., 2003; HALASA et al., 2007; HUIJPS et al., 2008; PIEPERS et al., 2009; FÁVERO et al., 2015; ROLLIN et al., 2015; VAN SOEST et al., 2016). The pathogenesis and incidence of this factor disease is influenced by the environment, the pathogen and the host (BURVENICH et al., 2003b). In addition to the socio-political importance, due to the impairment of animal welfare (BURVENICH et al., 2003b), the increasing use of antibiotics, combined with increasing resistances in bacterial populations (TEUBER, 2001; WALLMANN & HEBERER, 2014) is a major motivation for breeding robust dairy cattle with resistance to mastitis.

Increasingly, in dairy cattle breeding, not only phenotypic traits such as milkability, performance, fertility and milk somatic cell count are selected for, but genomic information is included (KUEHN et al., 2008). To learn more about the genetic background and relationship between molecular mechanisms and pathophysiological processes of susceptibility to udder infections, RNA-Seq analyses of the bovine udder transcriptome have been and are being performed.

The particularly frequent occurrence of clinical udder inflammation postpartum, in the period of early lactation, makes it necessary to sample and apply RNA-Seq at this time (BURVENICH et al., 2003b; VANGROENWEGHE et al., 2005; BURVENICH et al., 2007). However, up to 80% of the udder transcriptome during lactation (CANOVAS et al., 2014; IBEAGHA-AWEMU et al., 2016; BHAT et al., 2019) consists of transcripts from genes encoding casein (CSN1S1, CSN1S2, CSN2 and CSN3) and whey proteins (PAEP and LALBA), which makes it difficult to identify genes with lower expression levels. In order to avoid increased sequencing depth, combined with increased costs, and still allow for the precise identification of less strongly expressed transcripts that are relevant for udder regulation (BARTEL, 2004; HE & HANNON, 2004; LAWLESS et al., 2013; YANG et al., 2018), the proportion of highly abundant transcripts in the udder parenchyma transcriptome must be reduced.

Therefore, a method to deplete these highly abundant milk protein gene transcripts from the transcriptome of bovine udder parenchyma was developed and evaluated. In this study, the udder tissue of three Holstein-Friesian heifers, from one udder quarter with *E. coli* infection and one control quarter each was used. Oligonucleotides complementary to sequences near the 3'end of these transcripts were used to remove the highly expressed target transcripts. After hybridization of the DNA nucleotides to the target RNA a double-strand specific RNase H for digestion was used. As a result, the milk protein gene transcripts were deadenylated. In the subsequent step of poly(A) selection these transcripts were removed from the pool of total RNA and were no longer available for subsequent transcription in cDNA libraries and differential RNA-Seq analysis.

With this developed method we were able to reduce the proportion of milk protein gene transcripts in our samples from an average of 61% in udder quarters without infection to an average of 24%. Infected parenchyma showed an average of 30% strongly expressed transcripts, which could be reduced to an average of up to 9%. To show the improved sensitivity of RNA-Seq analysis for lowly expressed transcripts, we detected a higher proportion of significantly differentially expressed transcripts when comparing *E. coli* infected and uninfected udder quarters when applying the depletion method. Additionally, more unknown loci could be detected and the average FPKM values of the analyzed genes increased. Nevertheless, no unwanted bias in quantitative gene expression or in biological signal transduction pattern was observed.

The development and evaluation of a targeted depletion method to reduce the proportion of strongly expressed milk protein gene transcripts in the udder transcriptome was successful and can be applied in future studies to improve the sensitivity of RNA-Seq analysis for less highly transcribed genes. The accurate detection of regulatory relevant genes at a given sequencing depth thus facilitates the understanding of genetic background of the susceptibility of dairy cattle to udder infections. Most likely, a more detailed and cost-saving analysis of the transcriptome will help to identify phenotypic traits associated with mastitis susceptibility. It facilitates the identification and establishment of robust biomarkers that can support the breeding of mastitis resistant cattle.

9 Literatur

- Abebe R, Hatiya H, Abera M, Megersa B, Asmare K. Bovine mastitis: prevalence, risk factors and isolation of *Staphylococcus aureus* in dairy herds at Hawassa milk shed, South Ethiopia. *BMC Vet Res* 2016; 12: 270.
- Adiconis X, Borges-Rivera D, Satija R, DeLuca DS, Busby MA, Berlin AM, Sivachenko A, Thompson DA, Wysoker A, Fennell T, Gnirke A, Pochet N, Regev A, Levin JZ. Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nat Methods* 2013; 10: 623-9.
- Archer SK, Shirokikh NE, Preiss T. Selective and flexible depletion of problematic sequences from RNA-seq libraries at the cDNA stage. *BMC Genomics* 2014; 15
- Aronson JK, Ferner RE. Biomarkers-A General Review. *Curr Protoc Pharmacol* 2017; 76: 9 23 1-9 17.
- Barkema H, Schukken YH, Lam TJGM, Beiboer ML, Wilmink H, Benedictus G, Brand A. Incidence of Clinical Mastitis in Dairy Herds Grouped in Three Categories by Bulk Milk Somatic Cell Counts. *Journal of Dairy Science* 1998; 81: 411-9.
- Barkema HW, Green MJ, Bradley AJ, Zadoks RN. Invited review: The role of contagious disease in udder health. *J Dairy Sci* 2009; 92: 4717-29.
- Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 2004; 116: 281-97.
- Bayatti N, Cooper-Knock J, Bury JJ, Wyles M, Heath PR, Kirby J, Shaw PJ. Comparison of blood RNA extraction methods used for gene expression profiling in amyotrophic lateral sclerosis. *PLoS One* 2014; 9: e87508.
- Behjati S, Tarpey PS. What is next generation sequencing? *Arch Dis Child Educ Pract Ed* 2013; 98: 236-8.
- Bhat SA, Ahmad SM, Ibeagha-Awemu EM, Bhat BA, Dar MA, Mumtaz PT, Shah RA, Ganai NA. Comparative transcriptome analysis of mammary epithelial cells at different stages of lactation reveals wide differences in gene expression and pathways regulating milk synthesis between Jersey and Kashmiri cattle. *PLoS One* 2019; 14: e0211773.
- Brand B, Baes C, Mayer M, Reinsch N, Kuhn C. Identification of a two-marker-haplotype on *Bos taurus* autosome 18 associated with somatic cell score in German Holstein cattle. *BMC Genet* 2009; 10: 50.
- Buitenhuis B, Røntved CM, Edwards SM, Ingvarsten KL, Sørensen P. In depth analysis of genes and pathways of the mammary gland involved in the pathogenesis of bovine *Escherichia coli*- mastitis. *BMC Genomics* 2011; 12: 130.

- Burvenich C, Van Merris V, Mehrzad J, Diez-Fraile A, Duchateau L. Severity of *E. coli* mastitis is mainly determined by cow factors. *Vet Res* 2003; 34: 521-64.
- Burvenich C, Bannerman DD, Lippolis JD, Peelman L, Nonnecke BJ, Kehrli ME, Paape MJ. Cumulative physiological events influence the inflammatory response of the bovine udder to *Escherichia coli* infections during the transition period. *J Dairy Sci* 2007; 90
- Bush SJ, McCulloch MEB, Summers KM, Hume DA, Clark EL. Integration of quantitated expression estimates from polyA-selected and rRNA-depleted RNA-seq libraries. *BMC Bioinformatics* 2017; 18: 301.
- Cai Z, Dusza M, Guldbrandtsen B, Lund MS, Sahana G. Distinguishing pleiotropy from linked QTL between milk production traits and mastitis resistance in Nordic Holstein cattle. *Genet Sel Evol* 2020; 52: 19.
- Canovas A, Rincon G, Bevilacqua C, Islas-Trejo A, Brenaut P, Hovey RC, Boutinaud M, Morgenthaler C, VanKlompenberg MK, Martin P, Medrano JF. Comparison of five different RNA sources to examine the lactating bovine mammary gland transcriptome using RNA-Sequencing. *Sci Rep* 2014; 4: 5297.
- Castro TL, Seyffert N, Ramos RT, Barbosa S, Carvalho RD, Pinto AC, Carneiro AR, Silva WM, Pacheco LG, Downson C, Schneider MP, Miyoshi A, Azevedo V, Silva A. Ion Torrent-based transcriptional assessment of a *Corynebacterium pseudotuberculosis* equi strain reveals denaturing high-performance liquid chromatography a promising rRNA depletion method. *Microb Biotechnol* 2013; 6: 168-77.
- Cervinkova D, Vlkova H, Borodacova I, Makovcova J, Babak V, Lorencova A, Vrtkova I, Marosevic D, Jaglic Z. Prevalence of mastitis pathogens in milk from clinically healthy cows. *Veterinarni Medicina* 2013; 58: 567-75.
- Clark MB, Mercer TR, Bussotti G, Leonardi T, Haynes KR, Crawford J, Brunck ME, Cao KA, Thomas GP, Chen WY, Taft RJ, Nielsen LK, Enright AJ, Mattick JS, Dinger ME. Quantitative gene profiling of long noncoding RNAs with targeted RNA sequencing. *Nat Methods* 2015; 12: 339-42.
- De Vliegher S, Fox LK, Piepers S, McDougall S, Barkema HW. Invited review: Mastitis in dairy heifers: nature of the disease, potential impact, prevention, and control. *J Dairy Sci* 2012; 95: 1025-40.
- Demasius W, Weikard R, Hadlich F, Müller KE, Kühn C. Monitoring the immune response to vaccination with an inactivated vaccine associated to bovine neonatal pancytopenia by deep sequencing transcriptome analysis in cattle. *Veterinary Research* 2013; 44: 93.

- Detke S, Stein JL, Stein GS. Synthesis of histone messenger RNAs by RNA polymerase 11 in nuclei from S phase HeLa S3 cells. 1978;
- Do DN, Li R, Dudemaine PL, Ibeagha-Awemu EM. MicroRNA roles in signalling during lactation: an insight from differential expression, time course and pathway analyses of deep sequence data. *Scientific Reports* 2017; 7
- Esposti DD, Hernandez-Vargas H, Voegelé C, Fernandez-Jimenez N, Forey N, Bancel B, Le Calvez-Kelm F, McKay J, Merle P, Herceg Z. Identification of novel long non-coding RNAs deregulated in hepatocellular carcinoma using RNA-sequencing. *Oncotarget* 2016; 7: 31862-77.
- Fávero S, Portilho FVR, Oliveira ACR, Langoni H, Pantoja JCF. Factors associated with mastitis epidemiologic indexes, animal hygiene, and bulk milk bacterial concentrations in dairy herds housed on compost bedding. *Livest Sci* 2015; 181: 220-30.
- Freebern E, Santos DJA, Fang L, Jiang J, Parker Gaddis KL, Liu GE, VanRaden PM, Maltecca C, Cole JB, Ma L. GWAS and fine-mapping of livability and six disease traits in Holstein cattle. *BMC Genomics* 2020; 21: 41.
- Giannoukos G, Ciulla DM, Huang K, Haas BJ, Izard J, Levin JZ, Livny J, Earl AM, Gevers D, Ward DV, Nusbaum C, Birren BW, Gnirke A. Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome Biol* 2012; 13: R23.
- Gu HD, Das Gupta J, Schoenberg DR. The poly(A)-limiting element is a conserved cis-acting sequence that regulates poly(A) tail length on nuclear pre-mRNAs. *Proceedings of the National Academy of Sciences of the United States of America* 1999; 96: 8943-8.
- Gu W, Crawford ED, O'Donovan BD, Wilson MR, Chow ED, Retallack H, DeRisi JL. Depletion of Abundant Sequences by Hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. *Genome Biol* 2016; 17: 41.
- Gunther J, Esch K, Poschadel N, Petzl W, Zerbe H, Mitterhuemer S, Blum H, Seyfert HM. Comparative kinetics of *Escherichia coli*- and *Staphylococcus aureus*-specific activation of key immune pathways in mammary epithelial cells demonstrates that *S. aureus* elicits a delayed response dominated by interleukin-6 (IL-6) but not by IL-1A or tumor necrosis factor alpha. *Infect Immun* 2011; 79: 695-707.
- Gunther J, Petzl W, Zerbe H, Schuberth HJ, Seyfert HM. TLR ligands, but not modulators of histone modifiers, can induce the complex immune response pattern of endotoxin tolerance in mammary epithelial cells. *Innate Immun* 2017; 23: 155-64.

- Hagiwara S, Mori K, Okada H, Oikawa S, Nagahata H. Acute Escherichia coli Mastitis in Dairy Cattle: Diagnostic Parameters Associated with Poor Prognosis. *J Vet Med Sci* 2014; 76: 1431-6.
- Halasa T, Huijps K, Osteras O, Hogeveen H. Economic effects of bovine mastitis and mastitis management: a review. *Vet Q* 2007; 29: 18-31.
- Halvardson J, Zaghlool A, Feuk L. Exome RNA sequencing reveals rare and novel alternative transcripts. *Nucleic Acids Res* 2013; 41: e6.
- Hao Z, Zhou H, Hickford JGH, Gong H, Wang J, Hu J, Liu X, Li S, Zhao M, Luo Y. Transcriptome Profile Analysis of Mammary Gland Tissue from Two Breeds of Lactating Sheep. *Genes (Basel)* 2019; 10
- Harrington CA, Fei SS, Minnier J, Carbone L, Searles R, Davis BA, Ogle K, Planck SR, Rosenbaum JT, Choi D. RNA-Seq of human whole blood: Evaluation of globin RNA depletion on Ribo-Zero library method. *Sci Rep* 2020; 10: 6271.
- He L, Hannon GJ. MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet* 2004; 5: 522-31.
- Herbert ZT, Kershner JP, Butty VL, Thimmapuram J, Choudhari S, Alekseyev YO, Fan J, Podnar JW, Wilcox E, Gipson J, Gillaspay A, Jepsen K, BonDurant SS, Morris K, Berkeley M, LeClerc A, Simpson SD, Sommerville G, Grimmett L, Adams M, Levine SS. Cross-site comparison of ribosomal depletion kits for Illumina RNAseq library construction. *BMC Genomics* 2018; 19: 199.
- Hirakawa H, Oda Y, Phattarasukol S, Armour CD, Castle JC, Raymond CK, Lappala CR, Schaefer AL, Harwood CS, Greenberg EP. Activity of the Rhodopseudomonas palustris p-coumaroyl-homoserine lactone-responsive transcription factor RpaR. *J Bacteriol* 2011; 193: 2598-607.
- Hogan J, Larry Smith K. Coliform mastitis. *Vet Res* 2003; 34: 507-19.
- Holt RA, Jones SJ. The new paradigm of flow cell sequencing. *Genome Res* 2008; 18: 839-46.
- Huang Y, Sheth RU, Kaufman A, Wang HH. Scalable and cost-effective ribonuclease-based rRNA depletion for transcriptomics. *Nucleic Acids Res* 2020; 48: e20.
- Huijps K, Lam TJ, Hogeveen H. Costs of mastitis: facts and perception. *J Dairy Res* 2008; 75: 113-20.
- Ibeagha-Awemu EM, Li R, Ammah AA, Dudemaine PL, Bissonnette N, Benchaar C, Zhao X. Transcriptome adaptation of the bovine mammary gland to diets rich in unsaturated

- fatty acids shows greater impact of linseed oil over safflower oil on gene expression and metabolic pathways. *BMC Genomics* 2016; 17: 104.
- Jordan SC, Choi J, Kim I, Wu G, Toyoda M, Shin B, Vo A. Interleukin-6, A Cytokine Critical to Mediation of Inflammation, Autoimmunity and Allograft Rejection: Therapeutic Implications of IL-6 Receptor Blockade. *Transplantation* 2017; 101: 32-44.
- Kang EY, Martin LJ, Mangul S, Isvilanonda W, Zou J, Ben-David E, Han B, Lusis AJ, Shifman S, Eskin E. Discovering Single Nucleotide Polymorphisms Regulating Human Gene Expression Using Allele Specific Expression from RNA-seq Data. *Genetics* 2016; 204: 1057-64.
- Katinakis PK, Slater A, Burdon RH. Non-polyadenylated mRNAs from eukaryotes. *FEBS Lett* 1980; 116: 1-7.
- Krjutskov K, Koel M, Roost AM, Katayama S, Einarsdottir E, Jouhilahti EM, Soderhall C, Jaakma U, Plaas M, Vesterlund L, Lohi H, Salumets A, Kere J. Globin mRNA reduction for whole-blood transcriptome sequencing. *Sci Rep* 2016; 6: 31584.
- Kuehn C, Reinhardt F, Schwerin M. Marker assisted selection of heifers improved milk somatic cell count compared to selection on conventional pedigree breeding values. *Archiv fur Tierzucht* 2008; 51
- Kuhn C, Bennewitz J, Reinsch N, Xu N, Thomsen H, Looft C, Brockmann GA, Schwerin M, Weimann C, Hiendleder S, Erhardt G, Medjugorac I, Forster M, Brenig B, Reinhardt F, Reents R, Russ I, Averdunk G, Blumel J, Kalm E. Quantitative trait loci mapping of functional traits in the German Holstein cattle population. *J Dairy Sci* 2003; 86: 360-8.
- Kurz JP, Yang Z, Weiss RB, Wilson DJ, Rood KA, Liu GE, Wang Z. A genome-wide association study for mastitis resistance in phenotypically well-characterized Holstein dairy cattle using a selective genotyping approach. *Immunogenetics* 2019; 71: 35-47.
- Lawless N, Foroushani AB, McCabe MS, O'Farrelly C, Lynn DJ. Next generation sequencing reveals the expression of a unique miRNA profile in response to a gram-positive bacterial infection. *PLoS One* 2013; 8: e57543.
- Lee R, Feinbaum R, Ambros V. A short history of a short RNA. *Cell* 2004; 116: S89-92, 1 p following S6.
- Li L, Chen X, Chen Z. Identification of Key Candidate Genes in Dairy Cow in Response to *Escherichia coli* Mastitis by Bioinformatical Analysis. *Front Genet* 2019; 10: 1251.

- Li Z, Liu H, Jin X, Lo L, Liu J. Expression profiles of microRNAs from lactating and non-lactating bovine mammary glands and identification of miRNA related to lactation. *BMC Genomics* 2012; 13: 731.
- Lister R, Gregory BD, Ecker JR. Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. *Curr Opin Plant Biol* 2009; 12: 107-18.
- Mastrokolias A, den Dunnen JT, van Ommen GB, 't Hoen PA, van Roon-Mom WM. Increased sensitivity of next generation sequencing-based expression profiling after globin reduction in human blood RNA. *BMC Genomics* 2012; 13: 28.
- Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nature Reviews Genetics* 2009; 10: 155-9.
- Meredith BK, Berry DP, Kearney F, Finlay EK, Fahey AG, Bradley DG, Lynn DJ. A genome-wide association study for somatic cell score using the Illumina high-density bovine beadchip identifies several novel QTL potentially related to mastitis susceptibility. *Front Genet* 2013; 4: 229.
- Meyerholz M, Heimes A, Brodhagen J, Rohmeier L, Eickhoff T, Jander S, Huelsebusch A, Weikard R, Günther J, Seyfert H-M, Zerbe H, Petzl W, Hoedemaker M, Schuberth H-J, Engelmann S, Kuehn C (2018) ChronMast - a model to study functional genetic variation of mastitis susceptibility. *World Congress on Genetics Applied to Livestock Production Auckland*
- Mitterhuemer S, Petzl W, Krebs S, Mehne D, Klanner A, Wolf E, Zerbe H, Blum H. *Escherichia coli* infection induces distinct local and systemic transcriptome responses in the mammary gland. *BMC Genomics* 2010; 11: 138.
- Moore MJ, Proudfoot NJ. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* 2009; 136: 688-700.
- Niknafs YS, Han S, Ma T, Speers C, Zhang C, Wilder-Romans K, Iyer MK, Pitchiaya S, Malik R, Hosono Y, Prensner JR, Poliakov A, Singhal U, Xiao L, Kregel S, Siebenaler RF, Zhao SG, Uhl M, Gawronski A, Hayes DF, Pierce LJ, Cao X, Collins C, Backofen R, Sahinalp CS, Rae JM, Chinnaiyan AM, Feng FY. The lncRNA landscape of breast cancer reveals a role for DSCAM-AS1 in breast cancer progression. *Nat Commun* 2016; 7: 12791.
- O'Neil D, Glowatz H, Schlumpberger M. Ribosomal RNA depletion for efficient use of RNA-seq capacity. *Curr Protoc Mol Biol* 2013; Chapter 4: Unit 4 19.
- Ogorevc J, Kunej T, Razpet A, Dovc P. Database of cattle candidate genes and genetic markers for milk production and mastitis. *Anim Genet* 2009; 40: 832-51.

- Oliveira L, Hulland C, Ruegg PL. Characterization of clinical mastitis occurring in cows on 50 large dairy herds in Wisconsin. *Journal of Dairy Science* 2013; 96: 7538-49.
- Oliver SP, Sordillo LM. Udder health in the periparturient period. *J Dairy Sci* 1988; 71: 2584-606.
- Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet* 2011; 12: 87-98.
- Parker Gaddis KL, VanRaden PM, Cole JB, Norman HD, Nicolazzi E, Durr JW. Symposium review: Development, implementation, and perspectives of health evaluations in the United States. *J Dairy Sci* 2020; 103: 5354-65.
- Perkel JM. Visiting "noncodarnia". *Biotechniques* 2013; 54: 301, 3-4.
- Petzl W, Zerbe H, Gunther J, Yang W, Seyfert HM, Nurnberg G, Schuberth HJ. *Escherichia coli*, but not *Staphylococcus aureus* triggers an early increased expression of factors contributing to the innate immune defense in the udder of the cow. *Vet Res* 2008; 39: 18.
- Piepers S, De Vliegher S, de Kruif A, Opsomer G, Barkema HW. Impact of intramammary infections in dairy heifers on future udder health, milk production, and culling. *Vet Microbiol* 2009; 134: 113-20.
- Pol M, Ruegg PL. Treatment practices and quantification of antimicrobial drug usage in conventional and organic dairy farms in Wisconsin. *J Dairy Sci* 2007; 90: 249-61.
- Prezza G, Heckel T, Dietrich S, Homberger C, Westermann AJ, Vogel J. Improved bacterial RNA-seq by Cas9-based depletion of ribosomal RNA reads. *RNA* 2020; 26: 1069-78.
- Raboisson D, Ferchiou A, Pinior B, Gautier T, Sans P, Lhermie G. The Use of Meta-Analysis for the Measurement of Animal Disease Burden: Losses Due to Clinical Mastitis as an Example. *Front Vet Sci* 2020; 7: 149.
- Roe K, Gibot S, Verma S. Triggering receptor expressed on myeloid cells-1 (TREM-1): a new player in antiviral immunity? *Front Microbiol* 2014; 5: 627.
- Rollin E, Dhuyvetter KC, Overton MW. The cost of clinical mastitis in the first 30 days of lactation: An economic modeling tool. *Prev Vet Med* 2015; 122: 257-64.
- Ruegg PL. A 100-Year Review: Mastitis detection, management, and prevention. *J Dairy Sci* 2017; 100: 10381-97.
- Rupp R, Boichard D. Genetics of resistance to mastitis in dairy cattle. *Vet Res* 2003; 34: 671-88.

- Saini V, McClure JT, Leger D, Dufour S, Sheldon AG, Scholl DT, Barkema HW. Antimicrobial use on Canadian dairy farms. *J Dairy Sci* 2012; 95: 1209-21.
- Schulz J. Erkrankungen der Milchdrüse des Rindes, Grundsätze In: Euter- und Gesäugekrankheiten. Wendt K, Bostedt H, Mielke H, Fuchs H-W, eds. Jena-Stuttgart: Verlag Gustav Fischer 1994: 226-301.
- Seegers H, Fourichon C, Beaudeau F. Production effects related to mastitis and mastitis economics in dairy cattle herds. *Vet Res* 2003; 34: 475-91.
- Sender G, Korwin-Kossakowska A, Pawlik A, Hameed KGA, Oprzadek J. Genetic basis of mastitis resistance in dairy cattle - a review. *Annals of Animal Science* 2013; 13: 663-73.
- Shin H, Shannon CP, Fishbane N, Ruan J, Zhou M, Balshaw R, Wilson-McManus JE, Ng RT, McManus BM, Tebbutt SJ, Team PCoE. Variation in RNA-Seq transcriptome profiles of peripheral whole blood from healthy individuals with and without globin depletion. *PLoS One* 2014; 9: e91041.
- Sims D, Sudbery I, Illott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet* 2014; 15: 121-32.
- Sipka A, Klaessig S, Duhamel GE, Swinkels J, Rainard P, Schukken Y. Impact of intramammary treatment on gene expression profiles in bovine *Escherichia coli* mastitis. *PLoS One* 2014; 9: e85579.
- Staiger D, Brown JW. Alternative splicing at the intersection of biological timing, development, and stress responses. *Plant Cell* 2013; 25: 3640-56.
- Stein L. Genome annotation: from sequence to biology. *Nat Rev Genet* 2001; 2: 493-503.
- Suarez-Vega A, Gutierrez-Gil B, Klopp C, Tosser-Klopp G, Arranz JJ. Comprehensive RNA-Seq profiling to evaluate lactating sheep mammary gland transcriptome. *Sci Data* 2016; 3: 160051.
- Sultan M, Amstislavskiy V, Risch T, Schuette M, Dökel S, Ralser M, Balzereit D, Lehrach H, Yaspo M-L. Influence of RNA extraction methods and library selection schemes on RNA-seq data. *BMC Genomics* 2014; 15: 675.
- Teuber M. Veterinary use and antibiotic resistance. *Current Opinion in Microbiology* 2001; 4: 493-9.
- van Soest FJS, Santman-Berends I, Lam T, Hogeveen H. Failure and preventive costs of mastitis on Dutch dairy farms. *J Dairy Sci* 2016; 99: 8365-74.

- Vangroenweghe F, Lamote I, Burvenich C. Physiology of the periparturient period and its relation to severity of clinical mastitis. *Domest Anim Endocrinol* 2005; 29: 283-93.
- Wallmann J, Heberer T. Antibiotikaresistenzen: Eine ständig zunehmende Gefahr für Mensch und Tier und mögliche Auswege aus dem sich abzeichnenden Dilemma. *Journal für Verbraucherschutz und Lebensmittelsicherheit* 2014; 9: 1-2.
- Wang X, Ma P, Liu J, Zhang Q, Zhang Y, Ding X, Jiang L, Wang Y, Zhang Y, Sun D, Zhang S, Su G, Yu Y. Genome-wide association study in Chinese Holstein cows reveal two candidate genes for somatic cell score as an indicator for mastitis susceptibility. *BMC Genet* 2015; 16: 111.
- Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 2009; 10: 57-63.
- Wickramasinghe S, Rincon G, Islas-Trejo A, Medrano JF. Transcriptional profiling of bovine milk using RNA sequencing. *BMC Genomics* 2012; 13: 45.
- Wilhelm BT, Landry JR. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods* 2009; 48: 249-57.
- Wilusz JE, Freier SM, Spector DL. 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* 2008; 135: 919-32.
- Yang B, Jiao B, Ge W, Zhang X, Wang S, Zhao H, Wang X. Transcriptome sequencing to detect the potential role of long non-coding RNAs in bovine mammary gland during the dry and lactation period. *BMC Genomics* 2018; 19: 605.
- Yang D, Chen Q, Hoover DM, Staley P, Tucker KD, Lubkowski J, Oppenheim JJ. Many chemokines including CCL20/MIP-3 α display antimicrobial activity. *J Leuc Biol* 2003; 74: 448-55.
- Yang F, Liu LH, Li XP, Luo JY, Zhang Z, Yan ZT, Zhang SD, Li HS. Short communication: N-Acetylcysteine-mediated modulation of antibiotic susceptibility of bovine mastitis pathogens. *J Dairy Sci* 2016; 99: 4300-2.
- Yi H, Cho YJ, Won S, Lee JE, Jin Yu H, Kim S, Schroth GP, Luo S, Chun J. Duplex-specific nuclease efficiently removes rRNA for prokaryotic RNA-seq. *Nucleic Acids Res* 2011; 39: e140.
- Zabel BA, Rott A, Butcher EC. Leukocyte chemoattractant receptors in human disease pathogenesis. *Annu Rev Pathol* 2015; 10: 51-81.
- Zadoks RN, Middleton JR, McDougall S, Katholm J, Schukken YH. Molecular epidemiology of mastitis pathogens of dairy cattle and comparative relevance to humans. *J Mammary Gland Biol Neoplasia* 2011; 16: 357-72.

- Zhang XO, Yin QF, Chen LL, Yang L. Gene expression profiling of non-polyadenylated RNA-seq across species. *Genom Data* 2014; 2: 237-41.
- Zhao S, Zhang Y, Gordon W, Quan J, Xi H, Du S, von Schack D, Zhang B. Comparison of stranded and non-stranded RNA-seq transcriptome profiling and investigation of gene overlap. *BMC Genomics* 2015; 16: 675.
- Zhao W, He X, Hoadley KA, Parker JS, Hayes DN, Perou CM. Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC Genomics* 2014; 15: 419

10 Danksagung

Ich bedanke mich bei meiner Mentorin Prof. Dr. Christa Kühn für die Möglichkeit, an diesem spannenden Thema arbeiten zu können, für die hervorragende wissenschaftliche Betreuung und eine stets offene Tür bei Fragen oder Sorgen.

Bei Prof. Dr. Holm Zerbe bedanke ich mich für die Annahme dieser Arbeit an der Veterinärmedizinischen Fakultät der Ludwig-Maximilians-Universität München, für den stets freundlichen Kontakt bei den Projekttreffen, die kritische Durchsicht der Arbeit und die wertvollen Hinweise und Ratschläge, welche mir das Verfassen dieser Arbeit erleichtert haben.

Ich bedanke mich bei Frau Dr. Rosemarie Weikard dafür, an der Evaluierung dieser Methode arbeiten zu dürfen, für ihr immer offenes Ohr bei Fragen und für die Zusammenarbeit beim Schreiben der Publikation.

Ebenfalls geht mein Dank an alle Kooperationspartner des ChronMast-Projektes für die bereichernde Zusammenarbeit. Ganz besonders bedanke ich mich bei Frau Prof. Dr. Susanne Engelmann, Frau Prof. Dr. Martina Hoedemaker, Herrn Priv.-Doz. Dr. Wolfram Petzl, Herrn Prof. Dr. Hans-Joachim Schuberth, Frau Dr. Juliane Günther, Herrn Prof. Dr. Hans-Martin Seyfert sowie Frau Dr. Marie Meyerholz. Mein Dank geht an die Bundesanstalt für Landwirtschaft und Ernährung (BLE) und die Landwirtschaftliche Rentenbank für die Förderung dieses Projektes.

Des Weiteren möchte ich mich generell für die großartige Arbeit der aktuellen und ehemaligen Technischen Mitarbeiterinnen und Mitarbeiter der Abteilung Genomphysiologie am FBN bedanken. Ganz besonders bedanke ich mich bei Emma Schröder für das Einarbeiten in die Laborabläufe, bei Bärbel Pletz für die sorgfältige Isolation der RNA, bei Simone Wöhl für die wahrscheinlich gewissenhafteste Präparation der cDNA-Bibliotheken die man sich vorstellen kann, bei Sebastian Gaedecke für eine helfende Hand bei allen Abläufen und bei Manuela Lötze für ihre unendliche Geduld beim Pipettieren, Präparieren und Isolieren.

Ich bedanke mich ganz herzlich bei meinen ehemaligen Mitdoktorandinnen und Doktoranden, für die anregenden Gespräche und Hilfe in allen Lebenslagen. Ich danke Anne Berthold, Franziska Just und Zhiwei Li für die Motivation und positive Stimmung im Büro. Mein Dank geht an Wietje Nolte für ihre Hilfe beim Lösen kniffliger R-Probleme. Ich danke Ulrike Thom für ihre Hilfe im Labor und bei der Datenanalyse. Mein ganz besonderer Dank geht an meine ehemalige Kollegin und jetzt gute Freundin Annika Heimes, für ihre aufopfernde Hilfe in jeder Situation, ein immer offenes Ohr, gutes Essen und Aufmunterungen auch an schwierigen Tagen.

Außerdem geht ein besonderes Dankeschön an alle Mithelfenden bei den nächtlichen Autofahrten nach Hannover mit anschließender Probennahme.

Mein Dank geht an Frieder Hadlich für die bioinformatische Unterstützung und die motivierenden Gespräche.

Außerdem bedanke ich mich bei Frau Jana Masermann und Frau Rita Radloff für die schnelle Beantwortung organisatorischer Fragen.

Ich möchte mich bei meinen Freunden für die Unterstützung in den letzten Jahren bedanken. Ganz besonders bei meiner besten Freundin Aimée Neher, die trotz einer Entfernung von fast 1000 km immer für mich da ist.

Mein herzlichster Dank geht an meine Eltern Jens und Christina Hehl und meine Schwester Aileen Hehl für ihren Rückhalt und die liebevolle Unterstützung während dem Arbeiten an dieser Arbeit und während des Studiums.

Ich danke meinem Mann Ralf Brodhagen von ganzem Herzen für seine Geduld, seine Motivation in für mich scheinbar aussichtslosen Zeiten, seiner bedingungslosen Liebe und die Stütze, die er mir während der Zeit am FBN, aber auch während des Studiums war.

Zuletzt geht mein Dank noch an meinen geliebten Sohn Arthur. Sein Lachen hat mir über jede noch so schwierige Situation hinweggeholfen. Er hat mir beigebracht, die wirklich wichtigen Dinge im Leben zu erkennen.